DOCUMENT RESUME

ED 138 965 CS 003 481

AUTHOR Underwood, Benton J.; And Others
TITLE The Composition of Episodic Memory.
INSTITUTION Northwestern Univ., Evanston, Ill. Dept. of Psychology.
SPONS AGENCY Office of Naval Research, Arlington, Va. Personnel and Training Research Programs Office.
PUB DATE May 77
CONTRACT N00014-76-C-0270
NOTE 87p.

EDRS PRICE MF-$0.83 HC-$4.67 Plus Postage.
DESCRIPTORS *Associative Learning; Cognitive Processes; College Students; *Learning Processes; *Memory; *Reading Research; *Recall (Psychological); Retention; Semantics; Serial Learning; *Verbal Learning

ABSTRACT

This study examined the interrelationships among a number of episodic memory tasks and among various attributes of memory. A sample of 200 college students was tested for ten sessions; 28 different measures of episodic memory were obtained. In addition, five measures of semantic memory were available. Results indicated that episodic and semantic memory were generally unrelated. Among the measures of episodic learning, clustering was found to be unrelated to performance on other tasks. This was also true for a double-function, verbal-discrimination task and for a task designed specifically to measure susceptibility to interference. Twenty-two of the measures were included in a factor analysis from which five factors emerged. One factor was tied to free-recall tasks, another to paired-associate and serial tasks. Memory span constituted a third factor, a fourth factor involved involved verbal-discrimination lists, and a fifth was constituted by frequency assimilation and classical recognition. Reasons for the failure to form factors not tied to the type of task are discussed. (Author/AA)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

ED138965

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

Technical Report

THE COMPOSITION OF EPISODIC MEMORY

Benton J. Underwood, Robert F. Boruch, and Robert A. Malmi

May 1977

Northwestern University

Psychology Department

Sponsored by

Personnel & Training Research Programs
Psychological Sciences Division
Office of Naval Research
Arlington, Virginia
Contract No. N00014-76-C-0270

Approved for public release; distribution unlimited.

Reproduction in whole or in part is permitted

for any purpose of the United States Government.

CS003481

THE COMPOSITION OF EPISODIC MEMORY

Benton J. Underwood, Robert F. Boruch, and Robert A. Malmi

Northwestern University

We will report the results of a factor analysis of the scores on several of the tasks commonly used in our laboratories to study phenomena of learning and memory. Some of the reasons for undertaking the study, and some of the principles which guided the selection of tasks, will emerge in the background comments to follow.

Individual Differences and Attributes of Memory

Most of the tasks that were presented to our subjects were made up of common words. We assume that when subjects are asked to learn these tasks they produce, or abstract, certain kinds of information about a word and perhaps about its relationships with other words in the task. The information which is abstracted from word events is of central importance to this study, and we choose to speak of different types of information about words as being different attributes of memory (Underwood, 1969).

Attributes of the memories for words may reflect certain of the more or less physical aspects of the words, such as the orthography or the acoustic characteristics when pronounced. Of presumed greater importance for the performances of the young adult are the associative attributes, class attributes, imagery, and so on. When an attribute becomes a part of a memory, we speak of the encoding of the attribute. We should not, however, view encoding as being only a voluntary matter, although it may be such for certain types of information. The whole issue of encoding control and the utilization of attributes is confused and confusing at the present time. We had hoped to avoid

the issue but as will be seen, some of our findings refused to permit us to do so.

As indicated above, when a task made up of words is learned, the memory is presumed to consist of a certain group of attributes. Some of the attributes may not be of importance in determining immediate performance, some may be. Those that are most important in determining performance may be spoken of as the dominant attributes. To simplify the exposition, we will sometimes speak of a single dominant attribute as being the major determinant of performance. It is not unreasonable to assume that for optimal performance on a given task, a certain attribute should dominate. Such performance would be expected to occur when the dominant attribute for a learner is perfectly appropriate for the given task. Scores for the learning of a task inevitably show a broad distribution. In terms of the above conceptions, where do these individual differences originate? One source must be in the attributes which are allowed to dominate performance. Insofar as the attributes dominating performance deviate to varying degrees from the optimal, individual differences in performance will be present. A second source lies in the attributes per se. We believe that a distinction must be drawn between attributes available to the learner and those which dominate the performance. We have implied this above. However, all subjects may not abstract the same attributes from the same situation, and thereby a second source of individual differences arises. Later, a third source

will be pointed out, but for the moment we must hasten to say that our research was not intended to be analytical with regard to the two sources of individual differences identified. The question which we asked about individual differences was of a more general nature.

We assume that by the use of different tasks we can vary the attributes required for optimal performance. We were interested in the correlation of the performances across tasks because these correlations could tell us about the interrelationships among the memory attributes. Assume, for example, that we could show reliable individual differences among our learners in terms of the dominance of the imagery attribute, and that we could also show differences in the degree to which an acoustic attribute dominated performance. How do the individual differences on imagery dominance correlate with those for acoustic dominance? Or, do individual differences in the precision of the frequency attribute relate to individual differences in imagery utilization? Answers to such questions should influence the way in which we think about memory, and about the attributes which are presumed to constitute a memory.

Seeking answers to the above types of question was a major reason for undertaking the study. As we set about to provide answers, we faced a number of conceptual and measurement problems. Generally speaking, we do not have techniques whereby we can measure directly the "magnitude" or "amount" of an attribute. There are exceptions. One is the frequency attribute, which may be measured in a direct

sense by giving the learner events having varying frequencies and then requesting judgments of the frequencies. Normally however, we must infer the differences in an attribute from differences in learning scores. For example, differences in learning of abstract and concrete words may be said to represent differences in the operation of the imagery attribute. Is this a proper inference? We can only speak in terms of consent. Investigators who have studied the issue at great length, and who have manipulated many auxiliary variables, have reached the consensus that imagery differences are in some way accountable for the differences in learning which are found when abstract and concrete words are compared. As a general technique, then, the presence of certain attributes may be inferred from the manipulation of an independent variable.

It would seem rather straightforward to devise a measure of individual differences from such situations. Thus, in the case of imagery, we might presume that the degree to which a learner is facilitated by the concrete words (using performance on the abstract words as a base) could be used as a measure of individual differences on the imagery attribute. Some subjects will be facilitated a great deal (a large imagery attribute), others only a small amount if at all (small imagery attribute). Those who are aware of the inherent unreliability of difference scores will immediately conclude that this cannot be used as an appropriate measure of individual differences because the lack of reliability prohibits the possibility of

showing a relationship with other attributes (e.g., Cronbach & Furby, 1970). Suppose, for example, that we have two sets of scores on a group of subjects with the correlation between the two sets being 1.00. When the differences between the scores for individual subjects are examined it is apparent that they will all be of the same magnitude. A correlation involving such scores will, of course, be zero. As the correlation between two sets of scores decreases from 1.00, the differences scores will begin to show some variance, but correlations with another set of difference scores will necessarily be low. An illustration of the lack of correlation between difference scores will be given at a later point. There is a simple beauty in the logic for using difference scores as a means of identifying individual differences in attribute functioning; it is somewhat depressing that the beauty is destroyed by statistical fact.

What can be done about this problem? There is in fact no serious problem. We will illustrate this first by examining two tasks which might well appear in a factor analysis but which are not intended to measure a particular attribute. At the empirical level, we might well view a paired-associate list and a free-recall list as being sufficiently different so that they could be represented in different factors. To be indicative of different factors, of course, the direct correlation between the performances on the two tasks must be less than the reliabilities of either task. Now, we may treat a free-recall list of concrete words and a free-recall list of abstract words in exactly the same way. The abstract list has little possibility for use of the imagery attribute, the concrete list great

possibility. If there are wide individual differences in the use of the imagery attribute, the correlation between the performances on the two lists should be relatively low, and they may be shown to represent different factors.

How likely is the above chain of events? There is no a priori reason for concluding that it cannot happen, although one can certainly propose conditions which might prevent it from happening. One of these conditions must be discussed at this point. We indicated above that there was a third source of individual differences in performance, and it is appropriate to discuss this third source now. Many tasks, including free-recall tasks, are said to involve the formation of associations. We must expect that there will be individual differences in the rate at which new associations are formed. The use of the term "attributes" does not help our understanding of this; the formation of new associations fits into the attribute conception only in an indirect and insufficient manner. We could say that the formation of new associations occurs through the use of information present in semantic memory (by mneumonics of all varieties), and therefore that new associations are simply formed by the use of various attributes which we "summon" from semantic memory. Thus, individual differences in associative learning may be referred back to individual differences in attribute elicitation and utilization. Although we accept the notion that mneumonics may enter into the formation of allegedly new associations, the fact still remains that new associations must be capable of being formed without mneumonic aids, or we reach the in-

tolerable question of how associative aids were acquired in the first place.

Regardless of how associative formation occurs in a particular case, it is possible that the variance associated with individual differences in associative formation is so great that it will essentially mask or blot out the relatively small variance that might be associated with some of the attributes, e.g., imagery. Thus, the individual differences associated with the learning of a list of abstract words might be largely attributed to differences in associative learning, and this source of individual differences does not go away for a list of concrete words. The individual differences produced by the imagery attribute may be swamped by the individual differences in associative formation, differences which are independent of imagery.

If the above outcome is a possibility, we might attempt to devise situations in which the effect of the attribute could be magnified, and thereby reduce the relative effect of associative formation. In certain situations, an attribute will facilitate performance; the same attribute in other situations will retard performance. If subgroups of words in a serial list which belong to the same category (e.g., four animal names, four names of vegetables, and so on) are randomly distributed within a serial list, performance will be inhibited. The same words given in free recall will enhance performance. In the extreme case, this could produce a negative correlation in the performances on the two tasks providing the common factor

of associative-formation rate was not too dominant. Thus, the performance of subjects with a dominant conceptual associate attribute will be most enhanced in the free-recall task, most inhibited in the serial task. We included some situations of this type in our study.

The above is to indicate that the attempt to find the relationships among attributes of memory was faced with some potential difficulties. Even if we could not break through these difficulties we would not be without a substantial confusion because it could indicate to us that the attributes with which we have been concerned are not the dominant factors in the acquisition of new memories, at least where associative learning is fundamental to the mastery of the tasks.

There was one final consideration which seemed to us to provide a further leverage on the issue of attribute measurement. Attributes have been divided into two general classes (Underwood, 1969), those which are involved in the retrieval of memories (the associative or retrieval attributes) and those involved in discriminating among memories. This distinction refers to recall versus recognition. According to some theorists, retrieval attributes play no part in recognition performance. If this is true, then quite obviously we will expect recall and recognition tasks to fall into different factors even if we are unable to measure the relationships among particular attributes.

Choice of Tasks

The above discussion indicates that tasks were chosen with the

intent of measuring particular attributes of memory. We hoped to determine which attributes of memory "go together" and which ones do not. Generally speaking, in the process of selecting tasks that would be included for this purpose, we were at the same time selecting tasks which are frequently used as "standard" tasks in our laboratories. There is nothing mysterious about this conjunction. The effect of certain independent variables from which we have derived our ideas about attributes have been determined on frequently used tasks. If, then, we wanted to be most confident that we could produce an effect of a variable in our study, we would most assuredly use the tasks which have been used previously. Also, the fact that certain tasks are frequently used in the laboratory results from the fact that they achieve a certain end which, in some cases, means that they emphasize the operation of a particular attribute (in theory, at least). For example, we have believed that the verbal-discrimination task allows the frequency attribute to operate in reasonably pure form, so it is not unexpected that we included verbal-discrimination tasks in our study in addition to getting at the frequency attribute by measuring frequency discrimination directly.

So, we found that our attribute approach resulted in the choice of a certain number of tasks which have been used frequently in episodic memory research. But, there were other tasks which have been frequently used which did not get chosen by our attribute approach. To help reach other conclusions (see later), it seemed to us that we

should not be limited by the tasks which resulted from the attribute approach, and so we endeavored to include other tasks if we could within the general time limits we had imposed. For example, serial learning is the oldest task used in laboratory studies, but our attribute approach did not lead to this task. The serial task is one that has been particularly resistant to theoretical analysis by any approach, yet it was one that we were reluctant not to include in our study. Still further, there are tasks in which other investigators might have a substantial interest even if we did not. We felt some obligation to include these if we could work them in the schedule. The digit span is one such task. Two further factors were involved in choice of tasks. First, we found that in one case we had to construct a task to meet particular requirements. Second, we felt it desirable to get some information about semantic memory (in the Tulving sense, 1972) as a means of determining if performance measures of episodic memory were in any clear way associated with measures reflecting semantic memory.

By these considerations we developed or chose the tasks that we would use. Our data collection phase extended over a two-year period. It is the nature of a correlational study that once one gets well into the data-collection phase, it is essentially impossible (at least it is imprudent) to add tasks if one wishes to ever complete the data-collection phase with a substantial N. Yet, within a two-year period, advances in thinking and discoveries might strongly recommend adding new tasks or dropping some of those being administered. Being locked

into a plan, we could only mutter to ourselves about what we would have done could we start anew.

Some Implications

We have earlier indicated that we wished to try to understand how the attributes of memory are interrelated. While this desire may have been a sufficient reason for undertaking the research, we in fact had additional reasons of both a theoretical and practical nature which led us on. We will indicate these.

Assuming that a factor analysis would give us a small number of factors which were coherently related to tasks, it seemed to us that such information must necessarily be reflected in subsequent theorizing about the underlying mechanisms or processes. This may be illustrated with an issue (mentioned above) that has long been of concern to theorists, namely, the relationship between recognition and recall measures. If the tasks emphasizing recognition load on one factor and those emphasizing recall load on a different factor, it would seem that our theories about the memory processes must in some way reflect this. So too, if measures of memory span are unrelated to measures of serial learning, it would not seem appropriate to proceed as if the same one or two processes are fundamental in determining the performance on both tasks. At the same time, of course, when the performances on different tasks are shown to be related, it may advise us to develop our theories about these performances so that the commonality is reflected. All of this is to suggest that if the number of

factors which evolve are fewer in number than the tasks, it may help in constructing new theories or modifying old ones.

A second possible outcome, not unrelated to the first, has to do with empirical generalizations. Suppose that the outcome is a neat, ideal one in which three strong factors emerge. It is not unreasonable to conclude that all of the tasks falling within a factor would "react" comparably to the effect of a given independent variable. Thus, by examining experimentally the effect of a given independent variable on three key tasks (one from each factor), the generalization of the results may extend across all of the tasks involved in determining the factors originally.

A third implication has to do with the construction of a memory aptitude test. Again, assuming a relatively neat, definitive outcome in terms of factors, it should be possible to construct an aptitude test that could be administered in a relatively short period of time. Such a test would measure the abilities which underlie the performances on all of the tasks used in determining the factors. There would be numerous uses of such a test. It could be used in job placement if job analyses show the differential importance of certain types of memory factors. This test might be used in stratifying populations for subsequent experimental work. Such a test might also be used in diagnosing learning deficits on certain factors. If it can be shown that training would improve the performance of an individual on a particular factor, the test could be useful in determining

the remedial steps that could be taken.

Admitting the varying degrees of speculation involved in these implications, they were sufficiently compelling to recommend that we proceed with the study.

Previous Work

Other investigators have been concerned about the interrelationship among laboratory learning tasks but our searches failed to reveal any study which approached the scope we were contemplating. We will not, therefore, attempt any summary of the previous work at this time. In presenting our results, however, we will have both the opportunity and the need to look at our findings in conjunction with those of other investigators.

The Plan of Presentation

We will first describe the overall organization of the study. We will then describe each of the tasks which was used. In conjunction with the descriptions, we will also present some of the group data because (for some tasks) these data have intrinsic interest and value. As the third step we will present the correlation matrix followed by the results of factor analyses. As a fourth step we will discuss the implications of our findings.

Method

If our factor analysis was to be sensitive to differences of the relatively small magnitude expected by the attribute approach, it was essential that task reliability be substantial. Evidence available

indicated to us that this would not be a simple goal to achieve. When a task is given multiple trials, odd-even measures are normally highly correlated. For some purposes, this measure of within-task reliability is quite acceptable, but it is not an appropriate measure of reliability if the goal is to ask about relationships among tasks. The appropriate measure is the reliability between performances on two or more tasks of the same class which have been constructed so as to be as equivalent as possible. For example, 50 words might be sorted randomly to form two free-recall lists of 25 words each. The reliability of interest is indicated by the correlation between the performance on the two lists by the same subjects. Evidence we had available to us from work already done in our laboratory, and the reports of other investigators, indicated that demonstrating highly reliable performance between different tasks of the same class would not be expected. Our goal, therefore, was to plan the research so as to demonstrate merely substantial reliability.

Presumably, maximum stability in individual differences would be achieved for a given type or class of tasks by using many, many tasks from within the class. But, if the investigator also must use many types or classes of tasks, the element of time makes this approach unattractive, even impossible. The solution we chose was to use a practice task and a minimum of two additional tasks from a given class. The practice lists were always short versions of the regular or critical lists to follow, and always had the same charateristics

as the regular lists. Thus, if the critical lists involved categorized words, so also did the practice list.

Particular tasks, or variations within a larger task, were used to identify individual differences in encoding of a particular attribute. We set up an objective of using at least two different tasks to identify a particular attribute. We were far from successful in meeting this objective. In some cases this resulted from a failure of a variable to produce the same results as it had in previous experiment, and in other cases it was due to our inability to devise tasks to measure an attribute in more than one situation.

Within the constraints imposed by our attempts to obtain task reliability, and to measure attribute functioning reliably, the plan was developed to test subjects for 10 sessions of 50 minutes each during which 24 different tasks would be administered. A pilot study showed the plan was feasible, and gave us evidence on the adequacy of our mechanical techniques for presenting tasks and recording responses.

Because the great bulk of our experimental data on memory has been provided by college students, we used such students in the study. Any student answering the advertisement in the college newspaper was accepted, providing three criteria were met: (1) native language was English; (2) not a psycholgoy major; (3) not taking an introductory psychology course currently. The subjects were paid $25.00 for their participation. The testing of 100 subjects was completed during the 1974-75 school year, and an equal number was completed during 1975-76.

It was a remarkable coincidence that in both years the number of males was 43, the number of females, 57.

Subjects were tested in small groups, varying in size from 3 to 11. A given group was tested at the same hour on 10 days, Monday through Friday, of two consecutive weeks. All of the testing was conducted by one experimenter, a graduate student in psychology.

We will describe the tasks in an order that will emphasize common class names, such as paired-associate or free-recall tasks. This neither represents the order in which the tasks were presented to the subject nor the order which would necessarily emphasize theoretical relationships, but it will allow the theoretically neutral reader to use his past knowledge of tasks to a maximum advantage. Of course, the order of the tasks was constant for all subjects, an order that was largely dictated by time factors. As the tasks are described we will supply an abbreviation for each. A summary listing of all tasks will be given later (Table 2), and this table also indicates the day (1 - 10) on which each task was administered. As noted above, at least two tasks of each kind were given. Unless otherwise specified, a measure of reliability was determined by correlating the performance scores on the two tasks. In determining the relationship among the tasks, the mean of the scores on the two tasks of a given kind was used as the measure.

All of the items used for the learning tasks were words taken from various pools available to us. In many cases the words were

essentially a random sample from a still larger random sample taken from Thorndike-Lorge (1944). However, because particular words had to be used for some of the tasks in which associates and conceptual categories were involved, those particular words had to be eliminated from the pools before drawing randomly. Therefore, we can only describe the words used in some tasks as being essentially a random sample of a given class. No word was used more than once across tasks.

Visual presentation was used throughout via a projector controlled by a peripheral timer. The testing was carried out in a room used only for group experimentation.

Free Recall Control (FR-C)

The subjects were given four successive lists of 24 words each, presented for a single study and test trial. Each word was shown for 4 seconds for study, with 2 minutes allowed for recall. The words were five-letter words drawn from a pool consisting of all five-letter words given in the Thorndike-Lorge tables, and they were placed randomly into lists and into positions within the lists.

The means (number correctly recalled) for the four lists in order were 10.29 (3.12), 12.09 (3.68), 12.84 (3.13), and 13.17 (3.79). The standard deviations are in parenthesis, a convention we will use throughout. A practice effect is apparent. As a measure of reliability, the correlation between the sums of the number correct for lists 1 and 3 and the sums for lists 2 and 4 was used. The value was .68.

Recall as a function of position within the lists was examined. There were no systematic changes across lists and for the four lists combined, the recall-by-position curve was quite symmetrical. This is to say that primacy and recency effects were essentially equivalent.

Free Recall: Spacing (Fk-S)

We had not originally planned to include this task in the battery because the theoretical interpretation of the spacing effect was (and is) quite unclear. However, when we subsequently found that we could "squeeze it in" the schedule we decided to include it. Its inclusion expands the characteristics of the free-recall lists as a group, because the list is much longer than any of the other lists, and includes words presented twice within a single study trial.

Each list contained 32 words. Four of these occurred as primacy buffers, four as recency buffers. All of the remaining 24 words were presented twice, 12 as massed items (occurring twice in adjacent positions) and 12 as spaced items (the two occurrences being separated by at least three other items). The so-called lag for the spaced items varied from 3 to 20. The words were all of four letters. Each list was presented for a single study trial at a 2-second rate, with 160 seconds allowed for recall (5 seconds per word).

Overall mean recall was 13.66 (5.07) and 13.72 (4.73) for the two lists in order. The reliability (correlation between the overall recall values for the two lists) was .68, the same as for the free-

recall control lists described above. Across both lists, the mean recall for the 12 massed items was 3.55 (1.81), and for the spaced items, 6.08 (2.09), and the difference was reliable (t = 21.08). It was no surprise to find a marked superiority in the recall of the spaced items. The correlation between the recall of massed and spaced items was .69. With a reliability of .68 for the two lists, a correlation of .69 between the massed and spaced items means that we cannot possibly demonstrate that different attributes or different processes are involved in the acquisition of the two item types. So, we have used mean total correct responses for the two lists as the measure for between-task correlation, recognizing that we have no theory involved.

We had earlier mentioned that difference scores are inherently unreliable. This may be illustrated with the massed and spaced items. We noted that for the two lists combined, the correlation between the recall of massed and spaced items was .69. For the two lists separately, the values were .58 and .54. For each list, for each subject, the difference between the recall of spaced and massed items was determined, and these two distributions of 200 difference scores were correlated. The resulting value was .05.

Free Recall: Concrete (FR-CO) and Abstract (FR-AB)

Encoding by imagery is assumed to be an individual-differences variable. We hoped to evolve a measure of this variable by examining the free recall of concrete and abstract words. The Paivio, Yuille,

and Madigan (1968) norms were used to obtain words to construct two 24-item lists of concrete words and two corresponding lists of abstract words. The concrete words had values above six on the rating scale, the abstract words had values below three. The Thorndike-Lorge frequency was matched item by item for concrete and abstract words. The single study and recall trial for each list was carried out under exactly the same conditons as used for the free-recall control.

The means and standard deviations for the two lists of concrete words were 16.04 (4.37) and 14.66 (4.79), and the reliability was .70. For the two abstract lists the values were 11.15 (3.88) and 10.73 (3.76), with the reliability estimate being .60. Summing across both lists of each type showed a mean of 15.35 (4.22) for the concrete lists, 10.94 (3.42) for the abstract. Thus, a substantial difference (t = 19.17) resulted as would be expected from the great number of previous investigations dealing with this variable.

Free Recall: Interitem Associations (FR-II)

When common words are used in a learning task we assume that each word may elicit implicitly an associated word. These implicit associational responses (IARs) are assumed to be a part of the memory for the presented word. Under some circumstances, the IAR may facilitate acquisition, under others it may inhibit. In free recall, facilitation is expected. The term interitem associations as used here refers to any strong associate which is not a category name. The role

of conceptual IARs was examined separately in lists to be described later.

Each of the two lists consisted of 24 words made up of 12 pairs of associated words, e.g., doctor-nurse, shallow-deep. These associated pairs were obtained from a variety of sources along with the many others which were used in tasks to be described later. The 12 pairs were assigned randomly to the 24 positions within a list, subject only to the restriction that two associated words could not be contiguous. The procedures were exactly the same as for the free-recall control lists.

The means and standard deviations for the two lists were 18.72 (3.86) and 17.82 (3.92). The reliability was .69. Although we do not have an ideal control to determine whether or not the interitem associations facilitated acquisition, the means are distinctly higher than those for either the control lists or for the concrete and abstract lists.

It seemed obvious that we should obtain a measure of clustering for each subject and then determine how this relates to the performance on other tasks. We have done so. A simple clustering measure was used, namely, number of two-item clusters divided by total recall. Scores could vary from 0 to 50 (ignoring decimals). Correlating the values for the two lists provides a measure of reliability. The means for the two lists were 23.29 (15.32) and 29.71 (17.25), with the reliability estimate being .68. The distribution of scores

showed evidence of bimodality. Some subjects recalled in a serial manner exclusively, some recalled by associative clustering exclusively. The symbol for the clustering score for these lists is CL-II.

Free Recall: Conceptual Associations (FR-CA)

When a common word elicits implicitly the name of a category to which it belongs, we speak of it as a conceptual IAR. Each of the two 24-item lists was made up of three instances of each of eight categories. The items were selected from the tables prepared by Battig and Montague (1969). The 24 words were randomized in the list, subject only to the restriction that no words from the same category could be adjacent. The procedures were the same as for the free-recall control lists.

The means and standard deviations for the two lists were 16.61 (4.18) and 16.24 (4.29), with the reliability being .61. It seems likely that the presence of words from the same categories facilitated learning as has been true in so many past studies.

A clustering score for each subject was calculated as the number of adjacent recalls of items from a category divided by total recall. Ignoring decimals, scores could range from 0 to 67. The means for the two lists were 41.73 (20.30) and 39.28 (19.52), and the reliability estimate was .72. Characteristics of the distributions were much like those found for FR-II.

Paired-Associates: Control (PA-C)

All paired-associate lists contained 12 pairs, and learning was

by the study-test procedure for three trials. On the study trials, each pair was presented for 4 seconds, and on test trials, each stimulus term was shown for 4 seconds during which time the subjects wrote the appropriate response term if they could. Three different study orders were used in presenting the pairs, and three further orders of the stimulus terms were used during the test trials. These control lists consisted of five-letter words taken from the same pool as were the words for the FR-C lists. Pairing of the items was random with the restriction that the stimulus and response terms of a pair not have the same first letter.

The mean number of correct responses per trial was used as the response measure. These means and the standard deviations for the two lists in order were 9.37 (2.32) and 10.65 (1.46). Assuming the random assignment of words produced lists of equivalent difficulty, the means indicate a small practice effect which was reliable statistically ($t = 6.74$). The reliability was .75.

Paired Associates: Matching (PA-M)

The lists for this task were constructed in the same manner as were the control lists. The words were from the same pool and three study-test cycles were given. On the test trials, the stimulus terms were listed on the left side of the sheet, the response terms on the right. The subjects matched the stimulus and response terms by writing the appropriate response term after each stimulus term. The time allowed for matching was 60 seconds.

Data from our pilot subjects failed to reveal a problem which became apparent after the first 19 experimental subjects had been tested. The task was so easy that a few subjects obtained perfect performance after a single study trial. We viewed this task, of course, as a relatively pure measure of associative learning and considered it particularly important that the scores reflect reliable individual differences. Therefore, in an effort to "spread out" the learning, we reduced the study time per pair from 4 seconds to 2 seconds for all subsequent subjects. Although this had some effect in the expected direction, performance was still extremely high. For example, subjects 101-200 had mean total correct pairings of 32.42 and 32.51 on the two lists, out of a possible 36.

To determine reliability, it was necessary to use standard scores (x/σ). These scores for subjects 1-19 were based on the mean and standard deviation for those 19 subjects, and for the remaining 181 subjects, the standard scores were based on their mean and standard deviation. We were somewhat surprised to discover that the correlation between the scores for the two lists was .80.

Paired Associate: Crossed Associates (PA-II)

Each of these two lists was made up of 12 pairs of associated words (e.g., day-night, hammer-nail) inappropriately paired. It is known from previous work (e.g., Underwood & Ekstrand, 1968) that such lists result in interference in learning when the mean score on such a list is compared with the mean on a list without crossed

associates but otherwise comparable. Considering only the role of such associations, it would seem that a subject who is most facilitated in free-recall learning of lists with pairs of high associates (as described earlier), would be most interfered with in learning paired-associate lists made up of crossed associates. The procedure of testing for these lists was exactly the same as those used for the control lists.

The mean correct responses per trial were 10.56 (1.44) and 10.14 (1.81). The reliability was .77. The mean performance was roughly the same as found for PA-C. Although the control lists are not good controls (the words in the two types of lists may differ on a number of characteristics), it would certainly appear that no interference of consequence was produced by the crossed associates.

Paired Associates: Conceptual Interference (PA-CA)

These lists were designed to produce interference as a consequence of conceptual IARs. The 12 stimulus terms consisted of four instances of each of three categories. The response terms also consisted of four instances of each of three categories, but the categories were different from those used for the stimulus terms. All four instances of a category of the stimulus terms were paired with all four instances of a category of the response terms. Thus, four names for relatives (aunt, uncle, cousin, nephew) appeared as stimulus terms and were paired with the names of four fruits (peach, plum, apricot, grape), four names of metals were paired with four names of

alcoholic beverages, and four animal names were paired with the names of four tools. In a previous study (Underwood & Schulz, 1961), such lists were shown to produce substantial interference. The mechanism seems fairly obvious. The subjects learn immediately the manner in which the concepts are paired, e.g., relatives-fruits, metals-drinks, animals-tools. However, beyond this point in the learning, the category names are non-discriminative, and the interference occurs as the subjects try to learn the appropriate pairing within the concepts. The more persistent the implicit occurrence of the conceptual IARs, the greater the difficulty in learning. Thus, whereas the occurrence of conceptual IARs may facilitate free recall learning, they should inhibit the learning of these particular paired-associate lists.

The procedures were the same as for the control lists. For both study and test trials, the order of items was randomized, not blocked by categories. The mean number of correct responses per trial were 10.25 (1.51) and 9.99 (1.70), and the reliability was .67. Again, although we do not have a perfectly appropriate control, there is no evidence that interference was present.

Serial Learning: Control (SL-C)

The study-test method was used in presenting the 12-item serial lists. Each word was exposed for 2 seconds on the study trials, with 60 seconds allowed for the test trials. On the tests the subjects simply wrote as many words as possible in the correct position as indicated by a vertical column of 12 blanks. There were three study-

test cycles for each of the two lists. The words were five letter words from the pool referred to several times earlier.

Correct positioning was required for an item to be called correct. The mean total correct responses per trial were 9.03 (1.98) and 9.27 (2.06) for the two lists in order, with the reliability being .71. The usual bowed, serial position curve was evident, with position seven being of maximum difficulty.

Serial Learning: Positioning (SL-M)

The purpose of this task was to remove the recall requirement. On the test trials, the 12 five-letter words from the serial list were given in alphabetical order on the left side of the test page and the subjects' task was to write them in the appropriate blanks shown on the right. Study and test times were the same as for SL-C and again, three study-test cycles were administered.

The mean numbers of correct responses per trial were 9.67 (2.16) and 10.47 (1.84) for the two lists, and the reliability was .68. Performance on this task was only marginally better than performance on SL-C. Again, the usual serial-position effects were observed.

Verbal Discrimination: Control (VD-C)

In terms of our thinking, all of the tasks discussed thus far involved associative formation, and performance was fundamentally determined by associative or retrieval processes. The tasks we describe next are ones which emphasize the discrimination among memories, and in which associative retrieval processes may play only a minor role.

The verbal-discrimination lists consisted of 24 pairs, and were given for a single study and test trial. On the study trial, one word in each pair was underlined to designate to the subject the correct member of the pair. On the test trial, the underlining was omitted and the subject wrote the correct word for each pair. Each pair was presented for 2 seconds on the study trial, 4 seconds on the test trial. The order of the pairs on the test trial differed from that used on the study trial. Four lists were given. The words were a sample of two-syllable words with frequencies of from 1-10 in the Thorndike-Lorge (1944) tables.

The mean numbers of correct responses for the four lists were 21.70 (2.40), 20.72 (2.82) 20.79 (2.68), and 20.66 (2.96). To determine a reliability measure, the sums of the correct responses for Lists 1 and 3 were correlated with the sums for Lists 2 and 4. The r was .80.

Verbal Discrimination: Affective Cueing (VD-A)

We had wanted to include a task in which differences among subjects in the utilization of affective responses as discriminative cues could be detected. Biers (1970) had used a verbal-discrimination task in which all of the correct words had been rated high on the evaluative dimension of the semantic differential, and the incorrect words had been rated low. Very marked facilitation in learning was evident when compared with a control list of the same pairs in which correctness was inconsistently related to the level of the words on the evaluative dimension. Bier's finding led us to use this technique to

try to detect differences among subjects in their affective responses to words.

Two lists were used, one based on the evaluative dimension (List 1), the other on the potency dimension (List 2). The tables prepared by Heise (1965) were used as a source of words. The 24 correct words in List 1 had values of +1.00 or more on the evaluative scale, the incorrect words a value of -1.00 or less. The same cutoff values were used in choosing words for the potency dimension. It did not seem appropriate to us to use a practice list, nor to use two lists of a given type. However, both lists were given after the verbal-discrimination control lists so there should have been no problem of understanding the procedures, all of which were the same as for VD-C.

The mean numbers of correct responses were 20.62 (3.39) for List 1 (Evaluative) and 20.42 (3.67) for List 2 (Potency). It is obvious that if performance on the control lists is used as a reference, there is no evidence that the mean performance was influenced by the affective characteristic of the correct and incorrect words. The variability among the subjects seems to have been influenced, since the standard deviations were larger for these lists than for the control lists. One might suspect bimodality, but the frequency distributions gave no suggestion of it. The increase in variability seems to have been due entirely to a few subjects with poor performance. For the control lists, no subject got fewer than 13 correct responses; for List 1, 9 subjects scored lower than 13, and for List 2, 8 subjects

scored lower. The correlation between the performances on the two lists was .67. We view the results on these lists to be a distinct failure as far as getting at affective encoding. Just why Biers found such a large effect and we found none probably falls under a more general issue to which we will later address ourselves.

Verbal Discrimination: Double Functions (VD-DF)

In a double-function, verbal-discrimination list each item occurs in two different pairs. It is correct when a member of one pair, and incorrect when a member of the other. The learning of a double-function lists proceeds very slowly, and some subjects seem unable to make any headway. Viewed in the abstract, the double-function list could be mastered if a series of contingencies is acquired. Thus, the subject could learn that A is correct when paired with B, B is correct when paired with C, and so on. The evidence indicates that most subjects do not use this approach, or if they do, cannot apply it successfully (Underwood & Reichardt, 1975). Even when the number of contingency rules is small, learning is very slow. Essentially, the basis for the learning remains obscure.

Our original plans did not include this task in the battery because of the lack of theoretical understanding. However, when the pilot work showed that one of the other tasks was unreliable, we decided to include double-function lists. We might obtain evidence concerning the direction that theory development could take.

Each list consisted of 12 pairs, and the 12 words consisted of

four instances of each of three concepts (Battig & Montague, 1969). A list could be learned by mastering three contingencies. For example, in one of the lists there were four state names, four men's first names, and the names of four cities. When state names were paired with the names of cities, the state names were correct; when city names were paired with men's names, the city names were correct, and when men's names were paired with state names, the men's names were correct.

The lists were presented for four study and test trials using 4 seconds of study time for each pair and 4 seconds to respond to each pair on test trials. Eight different random orders of the pairs were used.

The mean numbers of correct responses per trial were 10.20 (1.89) and 10.35 (1.76). The reliability was .66. For the four trials for the two lists combined, the mean numbers correct were 9.33, 10.04, 10.79, and 10.92.

Running Recognition (RR-D)

In this task, developed by Shepard and Teightsoonian (1961), the subject was shown a long series of items, one at a time, and for each item the subject made a decision as to whether the item had or had not occurred earlier in the series. In our lists, we presented 174 words for List 1, and 173 words for List 2. Three measures were derived from each list.

1. A sensitivity measure of recognition consisting of the

sum of the false alarms on first occurrences of words and misses on the second occurrence of words.

2. A measure of false alarms to associates of words presented earlier in the series. This was presumed to be an index of the tendency of subjects to produce IARs to words.

3. A measure of the false alarms to the occurrence of the second member of pairs of homophones. This was designed to measure the dominance of the acoustic attribute in memory for the words.

The structure of the lists may now be described.

1. 20 critical stimulus words, each presented twice.
2. For each critical stimulus word there was a primary and secondary associate. For example, one critical stimulus word in List 1 was base. Subsequent to the second presentation of this word, the words ball (primary) and bottom (secondary) occurred. In some cases we used critical stimulus words which produced the same primaries or secondaries. In List 1 there were 16 different primaries and 18 different secondaries so that a total of 34 words was available in determining the false alarms produced by IARs. In List 2 there were 17 primaries and 16 secondaries, or a total of 33 words for determining false alarms produced by IARs.
3. 20 control words which occurred once and which were placed in positions equivalent to the positions held by the primary and secondary response words.

4. 20 pairs of homonyms, one word from each pair occurring in the first half of the list, the other in the second half of the list.
5. 20 neutral words which occurred twice.

The lists were presented at a 4-second rate. The slides were numbered and the subjects encircled YES or NO to indicate their response to each word. The measures used will now be described.

For the measure of regular recognition sensitivity or discrimination, which we will call D (hence, RR-D), we determined the number of false alarms on the first occurrence of 100 words, and the number of misses on the second occurrence of 40 words. The 100 words consisted of the first occurrence of 20 critical stimulus words, the first occurrence of the 20 neutral words occurring twice, the 20 control words, and the 40 homophones (see later). The 40 words for determining misses were the second occurrences of the 20 critical stimulus words and the 20 neutral words. The mean sums of the misses and false alarms were 12.58 (6.24) and 11.71 (7.65) for the two lists. The reliability was .70.

In determining the dominance of the acoustic attribute in memory for the words, we compared the false alarms on the 20 control items with the false alarms on the second occurring member of the 20 homophone pairs. If the acoustic attribute plays a role in memory, the number of false alarms should be greater for the words which had been preceded by a homophone than for the control words. There was no evidence that recognition was influenced by the homophones. For List 1, the mean false alarms were 2.62 (2.37) for the control items,

2.37 (1.90) for the homophones. The corresponding values for List 2 were 2.03 (2.61) and 2.21 (2.17). This attempt, therefore, to detect differences in acoustic encoding was not successful.

The dominance of the IARs in determining false alarms was assessed by comparing the number of false alarms made to the primary and secondary associates with the number made to control words, which included the second word of the homophone pairs (20 words) as well as the 20 words designated earlier as control words. There were 34 primaries and secondaries in List 1, 33 in List 2. The number of false alarms made to these items was incremented for each subject by a constant to a base of 40 to make these experimental items equivalent in number to the 40 control items.

For List 1, the mean number of false alarms to the experimental items was 7.46 (4.95), and to the control items, 5.96 (3.81). The correlation between numbers of control and experimental items was .79. The mean difference was highly reliable ($t = 7.14$). List 2 produced a different outcome. The mean for the experimental items was 4.42 (4.23), and for the control items, 4.26 (4.35), with a t of .76. The correlation was .77.

The above evidence indicates we observed a substantial IAR effect for List 1, but not for List 2. Because items were assigned to lists on a haphazard basis, there is no reason to believe that the different outcome for the two lists was determined by the particular items in the list. Rather, it appears that the subjects in some way learned to control the tendency to produce false alarms to the

associated items. Whatever the case, we must conclude that our attempt to obtain reliable individual differences on the associative attribute as manifested in the performance on the running-recognition task was unsuccessful. Furthermore, the substantial correlations between the number of errors on the experimental and control items indicate that very likely the false alarms to both types of items are produced by the same underlying mechanisms.

Situational Frequency (SF-Z)

In this task the subject made judgments of the frequency with which words occurred in a long list. There were 92 positions in each list resulting from 12 words presented once, 12 presented twice, 12 presented three times, and four presented five times. The items were presented at a 2-second rate. The tests were unpaced. The 40 words occurring in the study list were randomized along with 12 new words and the subject made an absolute judgment of the frequency of each of the 52 words. The words were all of two syllables and had Thorndike-Lorge (1944) frequencies falling between 1 and 10.

The mean judged frequencies were linearly related to true frequency. As a measure of individual performance we used the correlation between true frequencies for the 52 words and the judged frequencies. The mean correlations were .87 and .85 for Lists 1 and 2. For statistical purposes, the correlations were transformed to z'. The mean z' values were 1.32 (.30) and 1.26 (.40), and the reliability was .67.

List Differentiation (LD)

This task represents one of those used to measure individual differences in the ability to order memories appropriately on the time dimension. We speak of this as the temporal attribute. The subject was given three successive lists of 20 four-letter words each. Each word was presented for 2 seconds. The subject was clearly informed when one list was completed and when another list began. For the unpaced test the 60 words were randomized, and the subject circled one of three numbers (1, 2 or 3) to indicate list membership. The subjects were required to respond to all 60 words, guessing if necessary.

The mean total errors was used as the response measure. The values were 28.82 (8.40) and 24.99 (9.72) for the two sets of lists in order; the reliability was .71.

Some other comments about the findings may be appropriate. The mean numbers of errors made on the three lists in order were 8.47, 8.88, and 9.55. Of the errors made on the first of the three lists, 66% were identified as having been in the second list, and 34% were identified as having been in the third list. Errors made on the second list were divided 58% and 42% between the first and third. Errors on the third list were split about equally between the first and second lists.

We had planned originally to have a second test of the temporal attribute. In this test, the subjects were given a series of

relatively short lists. On the tests given after each list, pairs of words from the lists were displayed to the subjects and they were asked to make recency judgments (which word occurred most recently in the list) and lag judgments (how many other words fell between the two test pairs). Our pilot work showed these performance measures to be unreliable, and the test was dropped from the final battery.

Memory Span: Digits (MS-D)

The strings were 6, 7, 8 and 9 digits in length, eight strings of each. These were preceded by six strings of five digits each, used as practice. Each string was produced randomly using all numbers 1 through 9, except that: (1) no more than two numbers were allowed in natural sequence, either forward or backward; (2) no adjacent strings were allowed to end with the same last number nor to start with the same number. The method of complete presentation was used, with the exposure duration being .5 seconds per digit. One second per item was allowed to recall each string. The responses were written on a prepared sheet in which the appropriate number of blanks was present for each string length. All eight strings of six digits were given before the eight strings of seven digits, and so on, from shortest to longest.

Two scoring techniques were used, namely, number of correct strings and number of digits correct per string. These two measures correlated .95, so we have used number of letters correct per string since the reliability (based on odd-even strings) was a little higher (.75). Performance on strings of six digits was almost perfect for

all subjects. The percentages of errors made at each of the positions for string lengths of 7, 8 and 9 digits are shown in Figure 1. The most noteworthy fact is that at all three lengths the maximum numbers of errors were made on the next to the last number in the series, regardless of length. This finding replicates one reported by Chiang and Atkinson (1976).

Memory Span: (Letters MS-LL and MS-HL)

The memory span for letters was used as a second index of the dominance of the acoustic attribute. One set of strings was made of letters with high acoustic similarity, and one of letters of low acoustic similarity, as given by Conrad (1964). The nine letters used for the high-similarity strings were B, C, D, E, G, P, T, U, V (MS-HL). The low similarity strings were: B, H, J, L, O, K, R, W, Y (MS-LL). The low-similarity strings were given prior to the high-similarity strings. There were six practice strings of five letters each, followed by four strings each of 6, 7, 8 and 9 letters. Construction of the strings was handled by assigning letters on a random basis subject only to the restriction that successive strings not have the same first letter or the same last letter. All other procedures were the same as for the digit spans.

The odd-even reliability (numbers of letters correct) for the low-similarity strings was .64; that for the high-similarity strings, .71. The number of errors as a function of position was much the same

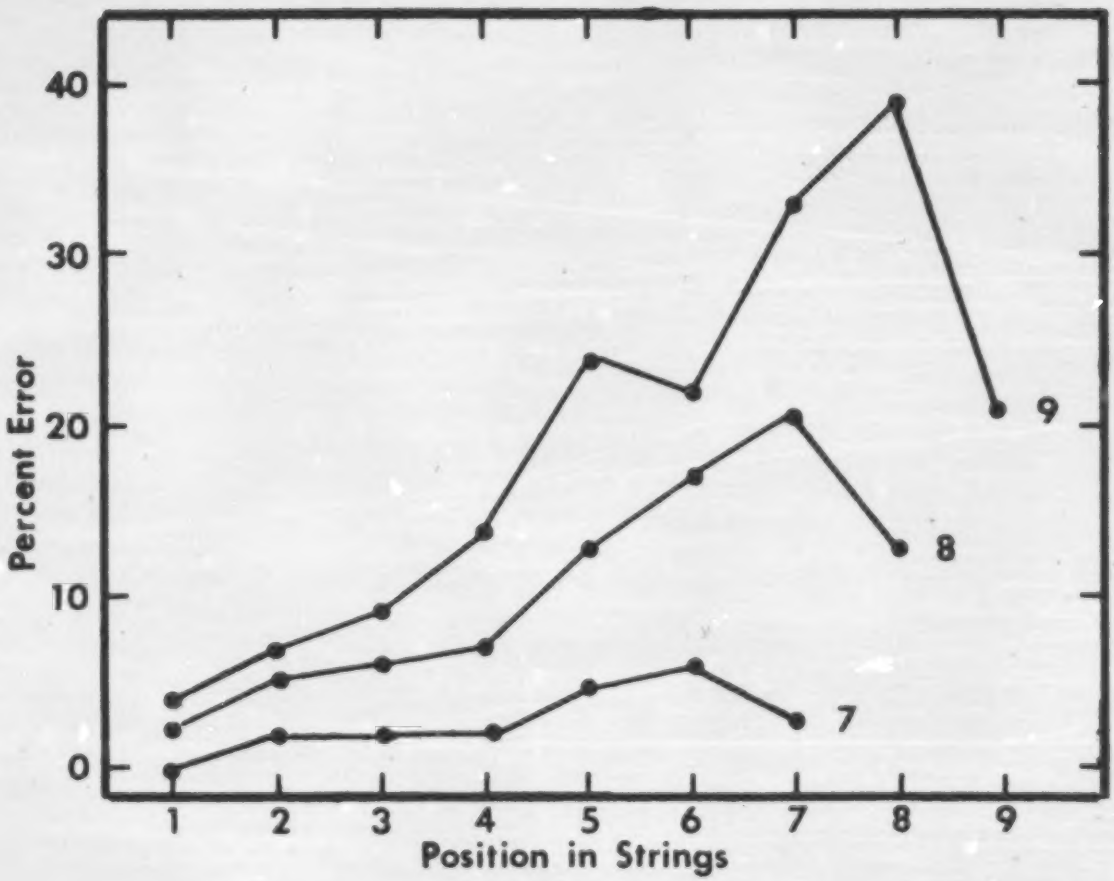

Figure 1. Percent errors as a function of position in string and string length (7, 8, 9) for digit span.

as shown in Figure 1 for the digits. Figure 2 shows the percent correct letters as a function of string length and similarity. Although the difference between the high- and low-similarity letters was reliable (t = 3.91), we were somewhat disappointed by the small magnitude of the acoustic-confusion effect.

Interference Susceptibility (IS)

We had wanted a task to measure individual differences in susceptibility to interference by associations established within the laboratory. The task finally used consisted of a series of short (five pairs) paired-associate lists, having three-letter words as stimulus terms, and the numbers 1 - 5 as the response terms. These numbers were used for all lists. The use of a closed system of numbers as response terms eliminated the necessity of acquiring responses per se, and made the primary task that of establishing associations between stimulus and response terms.

There were six sets of paired-associate lists, each set consisting of four lists. The four lists within a set differed only in the pairing of the numbers with the stimulus terms, the same five stimulus terms being used for all four lists within a set. Thus, within a set, the A-Br paradigm held across the four lists. The six sets differed only in that a different group of five, three-letter words was used as stimulus terms. Therefore, across sets, the C-B

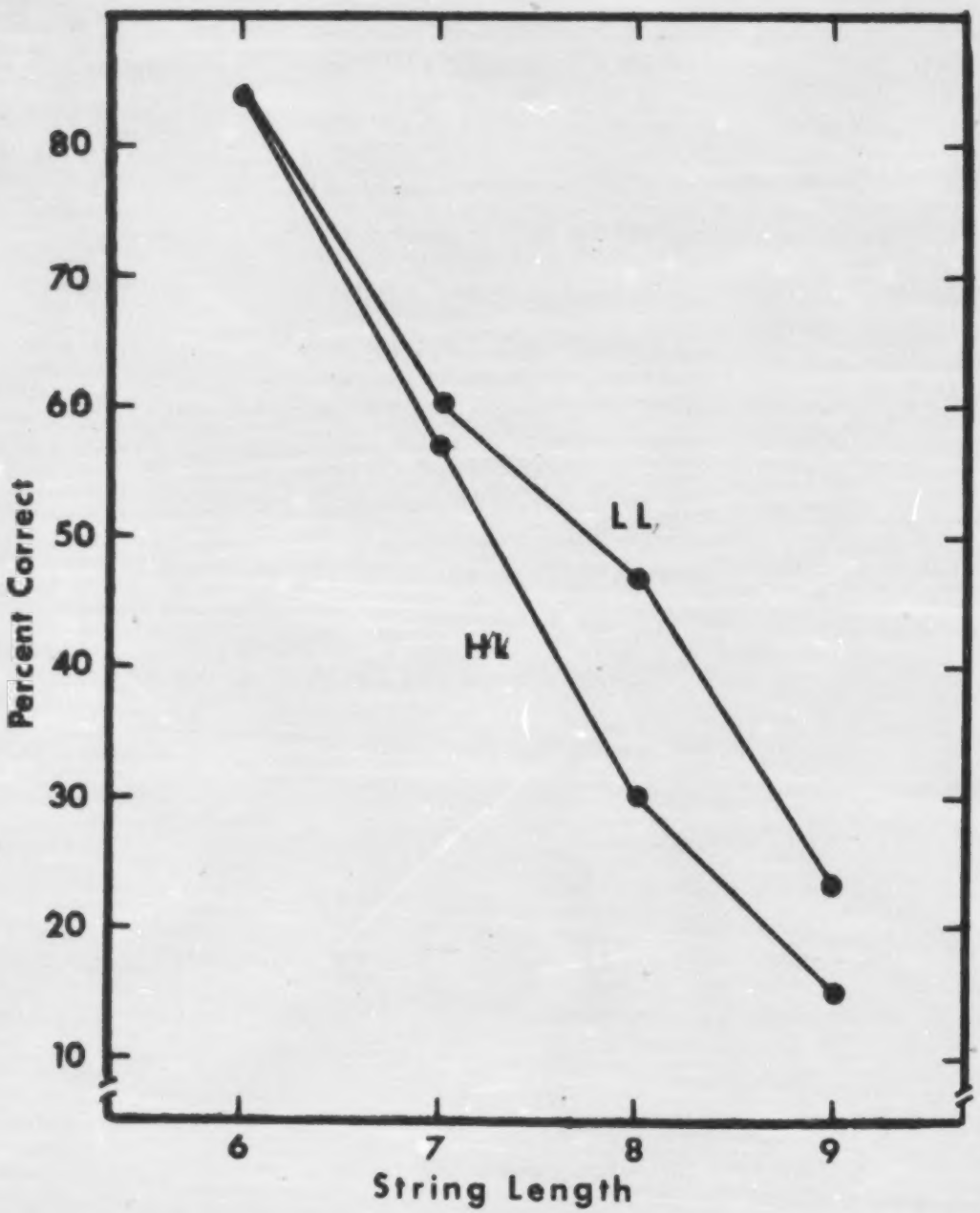

Figure 2. Percent correct in letter spans as a function of string length and low (LL) and high (HL) letter confusability.

paradigm held. We anticipated that performance would decrease within each set, and also decrease across sets. The subject was given a single study and test trial on each of the 24 lists. The study and test trials were conducted at a 3-second rate.

The performance on these lists was highly reliable. The sum of the correct responses for sets 1, 3 and 5 correlated .81 with the number of correct responses for sets 2, 4 and 6. The same value was found when the sum of lists 1 and 3 across sets was correlated with the sum of lists 2 and 4. As may be seen in Figure 3, the expectations concerning interference were partially borne out. In all sets, performance decreased across the four lists. Across sets, performance actually increased for the first four sets, with a decline on the fifth and sixth sets.

The fact that task reliability was high did not mean that we had a reliable measure of individual differences in susceptibility to interference. Given the results as shown in Figure 3, it would seem that the most appropriate measure would be a slope measure which would indicate a decline in performance across lists within sets. Such a slope measure could neutralize differences in learning ability per se as manifested on the first lists within sets. We derived several such measures but in all cases the correlations with total correct responses were so high that it would simply not be possible

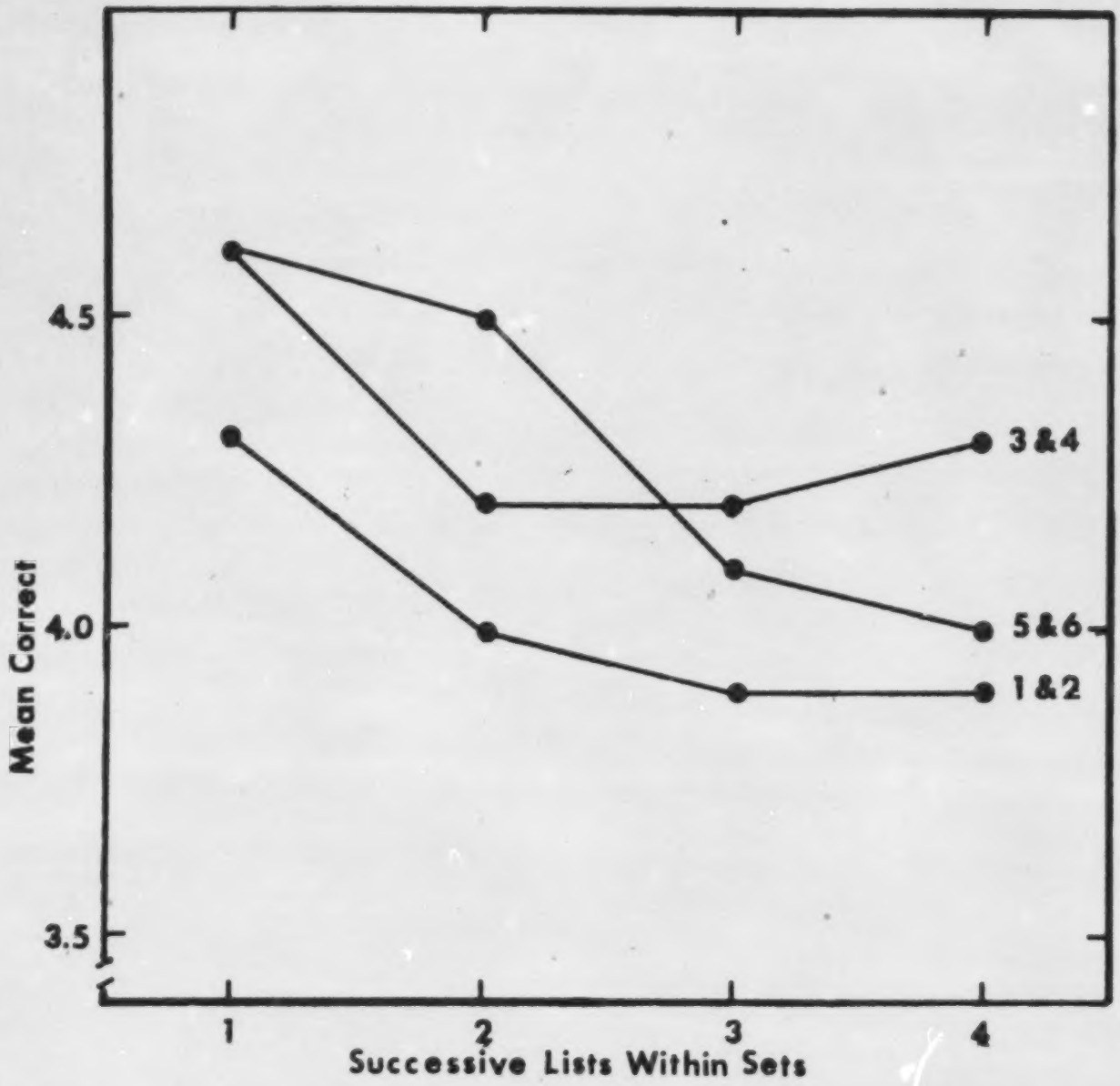

Figure 3. Mean correct responses as a function of lists within sets and sets (1 and 2 combined; 3 and 4 combined; 5 and 6 combined). This test was designed to measure susceptibility to accumulative interference (Task IS).

for the two measures to load on different factors. Therefore, we have retained only the total correct measure.

Simultaneous Tasks (SA)

We come now to the final learning task, which actually consisted of several tasks, all learned simultaneously. There were three reasons for including this task. First, we have been using the task extensively in experimental research and we had a strong interest in its relationship to the many other tasks used in the present study. Second, the subtasks were known to be relatively independent. Third, we would further extend the generality of our findings based on types of tasks, e.g., free recall, recognition.

In presenting the tasks, the subjects were fully informed of the nature of the material they would be shown, and the nature of the tests over the material. The subjects were asked to imagine that they were driving through an urban area. On the slides which they were shown was verbal material of the kind that they would, in fact, see on such a trip. Those materials were described, along with the method of testing.

Company Names (SA-FR). There were 30 different fictitious company names, all consisting of two words, with the second word indicating a product or service, e.g., Eagle Chemicals. Each pair occurred twice. The subjects knew that the company names had to be recalled on the test, hence, it was a free-recall test (SA-FR).

Traffic and Points of Interest Signs (SA-D). There were 32

such signs, all consisting of two words, e.g., EXIT RAMP, MEMORIAL CEMETERY. Twelve of the signs were shown once. For 12 further signs, each of the two words of a given sign appeared on different slides. The subjects knew that on the test they would have to make YES-NO decisions as to which signs were presented as intact signs (the two words presented together on a slide). The test consisted of the 24 two-word signs as noted above, plus eight new signs. Presenting the two words separately for some signs is of no moment for the present study. Rather, we view this task as a simple recognition task (SA-D) and the response measure used was the number of misses plus the number of false alarms (to new signs and to those presented as separate words on the study trials).

State Names (SA-Z). Names of 10 states, presumed to simulate names seen on license plates, occurred with varying frequencies. Two state names occurred with each of the following frequencies: 1, 3, 6, 10, 15. On the test, the subjects were given the 10 state names plus two new names and they made absolute frequency judgments for each of the 12 names. The measure of the precision of frequency assimilation for each subject was the correlation between true and judged frequency, with the correlations transformed to z'.

Street Names (SA-O). Seven different street names appeared during the drive, relatively evenly spaced throughout the series of slides. On the test the subjects were given the seven street names and were asked to assign the numbers 1 through 7 to them to indicate the order (O) in which the streets were crossed. We thought of this

as a measure of temporal discrimination. The response measure was the correlation between true order and judged order.

All of the material was presented on 24 slides. Of course, several different words occurred on each slide, and in our thinking, the memory was flooded with information to be learned. Each slide was presented for 10 seconds. The four tasks (recall, recognition, frequency, temporal) were always tested in the same order, but we know from other data that there was no loss over the time required to conduct the tests. After the first set of tests, the slides were presented a second time in the same order as on the first trial, and again the tests were given.

For this task we must accept as measures of reliability the correlations between performance on the two trials. The mean performance on each trial, their standard deviations, and the correlations are given in Table 1. Two points should be made about the data in Table 1. First the reliabilities are judged acceptable except for the ordering task. As noted before, performance on judgments of temporal ordering using other tasks has been low. Nevertheless, we have retained this variable in some subsequent analysis on the grounds that the sum of the two trials may have sufficient reliability to show relationships if such exist. Second, it will be observed that performance on the second trial of frequency judgments is a little lower than was found for the first trial. This has been found in several

Table 1

Performance on the Simultaneous Tasks and the Reliability of These Tasks as Measured by the Trial 1 x Trial 2 Correlations

(The measure for recall was mean correct; for frequency judgments, the z' transformation of the r between true and judged frequency; for recognition, the sum of the false alarms and misses; and for ordering, the z' transformation of the correlation between true and judged ordering.)

	Trial 1		Trial 2		reliability
	M	σ	M	σ	
Recall	1.86	1.59	6.91	3.41	.64
Frequency	1.30	.50	1.23	.49	.59
Recognition	7.68	3.42	5.02	3.32	.58
Ordering	.46	.58	.70	.65	.23

other cases (as yet unpublished); the reason for the failure of performance to improve is obscure.

Background Frequency (BF)

We wanted to get some measure of differences among individuals in semantic memory in order to determine if semantic memory exerted any control over episodic memory. One of these tasks dealt with the background frequency of words. We chose 103 pairs of words. The two words in a pair had the same number of syllables, were the same parts of speech, and were judged to be roughtly equivalent on a concrete-abstract dimension. Words which in our judgment might have had gross frequency changes in recent decades (e.g., jet) were not used. The two words in each pair differed in frequency as determined by the Thorndike and Lorge (1944) tables. The subjects' task was to choose the word in each pair having the highest frequency. The frequency difference varied across pairs as did the base frequency.

This was an unpaced test. The 103 pairs were randomized in four columns on two pages, and the subjects were instructed to choose the word in each pair which was used most frequently in magazines, books, and newspaper. A decision was required for each pair. The mean number of errors was 36.49 (5.66). The reliability, calculated as the correlation between the total errors in column 1 plus 3 and 2 plus 4, was .28. We had not expected a problem of reliability with this task. As will be seen, the scores for this test failed to correlate with scores on other tasks and we are left in the ambiguous posi-

tion of not knowing whether the failure to correlate was due to a true lack of relationship with other tasks or to a lack of reliability of the test scores on the judgments of background frequency.

Vocabulary (V)

As a vocabulary measure we used a test developed earlier (Zimmerman, Broder, Shaughnessy, & Underwood, 1977). For this test the subject was presented 76 words of varying frequencies, and 24 non-words. These non-words were made by putting together syllables from real words so that they were relatively easy to pronounce. For each of the 100 units the subject made a rating along a six point scale as to his degree of confidence that the unit was a word. The measure of vocabulary knowledge was the difference between the mean ratings for the words and non-words divided by the pooled variance of the ratings. The mean score for the 200 subjects was .66 (.28). There were two pages in the test booklet. As a measure of reliability, we correlated the scores for page 1 with those for page 2. The value was .61.

Spelling (S)

We constructed a spelling test along the lines used to construct the vocabulary test. From various sources we obtained 100 words which are commonly misspelled. On a random basis, 50 correct spellings were chosen, and 50 incorrect spellings. These were randomized in a two-page test booklet. The subjects rated each unit as to their degree of confidence that it was spelled correctly. Spelling ability

was measured in the same way as was vocabulary ability, and the mean was .83 (.46). The reliability (.67) was determined by correlating the score for the items on page 1 with those on page 2.

Finally, it should be mentioned that we were able to find verbal and mathematical scores from the SAT for 194 of the 200 subjects. The mean for the verbal scores (SAT-V) was 59.99 (6.63), for the math (SAT-M) scores, 62.28 (8.57).

Results

Intercorrelations

Some of the distributions of raw scores were negatively skewed. Cubed scores ameliorated the skewness, but the transformed values produced no appreciable and consistent differences in the correlations. The data with which we will deal are based on nontransformed scores.

The 33 variables which were used to produce the correlation matrix are summarized in Table 2. A few of the response measures were error measures, whereas most of them were correct-response measures. In presenting the correlations, the negative signs were removed when a positive relationship was intended. Thus, any negative correlation in Table 3 represents a true negative relationship. The decimal points for the correlations are omitted.

Even a quick perusal of Table 3 will show two facts. First,

Table 2

The 33 Variables Producing the Intercorrelation Matrix

(The day on which each task was administered is given in the second column, the reliability of each task in the third column.)

Variable Number	Day	Reliability	Abbreviation	Description
1	1	.68	FR-C	Free recall, control
2	4	.68	FR-S	Free recall, spaced and massed
3	6	.70	FR-CO	Free recall, concrete words
4	6	.60	FR-AB	Free recall, abstract words
5	8	.69	FR-II	Free recall, pairs of associates
6	8	.61	FR-CA	Free recall, categorized lists
7	8	.68	CL-II	Clustering, in FR-II
8	8	.72	CL-CA	Clustering, in FR-CA
9	4	.75	PA-C	Paired associates, control
10	5	.80	PA-M	Paired associates, matching
11	10	.77	PA-II	Paired associates, crossed associates
12	10	.67	PA-CA	Paired associates, paired categories
13	6	.71	SL-C	Serial learning, control
14	8	.68	SL-M	Serial learning, matching
15	5	.80	VD-C	Verbal discrimination, control
16	5	.67	VD-A	Verbal discrimination, affective words
17	9	.66	VD-DF	Verbal discrimination, double function
18	2	.70	RR-D	Running recognition, basic recognition measure
19	3	.67	SF-Z	Situational frequency judgments
20	2	.71	LD	List discrimination
21	3	.75	MS-D	Memory span, digits
22	3	.64	MS-LL	Memory span, low-similarity letters
23	3	.71	MS-HL	Memory span, high-similarity letters
24	7	.81	IS	Interference susceptibility
25	7	.64	SA-FR	Simultaneous acquisition, free-recall of pairs
26	7	.59	SA-Z	Simultaneous acquisition, frequency discrimination
27	7	.58	SA-D	Simultaneous acquisition, recognition of pairs
28	7	.23	SA-O	Simultaneous acquisition, ordering
29	1	.28	BF	Background frequency judgments
30	9	.61	V	Vocabulary test
31	4	.67	S	Spelling test
32	---	---	SAT-V	Scholastic aptitude test, verbal
33	---	---	SAT-M	Scholastic aptitude test, mathematics

Table 3

Intercorrelations Among the 33 Variables for 200 Subjects

(decimals omitted)

		1 FR-C	2 FR-S	3 FR-CO	4 FR-AB	5 FR-II	6 FR-CA	7 CL-II	8 CL-CA
1.	FR-C	--	66	59	56	56	49	-7	-5
2.	FR-S	66	--	65	54	56	53	9	9
3.	FR-CO	59	65	--	66	62	54	1	2
4.	FR-AB	56	54	66	--	59	53	3	-1
5.	FR-II	56	56	62	59	--	72	28	23
6.	FR-CA	49	53	54	53	72	--	20	37
7.	CL-II	-7	9	1	3	28	20	--	71
8.	CL-CA	-5	9	2	-1	23	37	71	--
9.	PA-C	53	53	45	44	48	40	13	9
10.	PA-M	42	47	43	42	43	47	7	18
11.	PA-II	48	46	47	45	50	46	9	13
12.	PA-CA	44	41	44	47	51	44	11	13
13.	SL-C	41	51	41	43	40	39	-9	-4
14.	SL-M	48	46	52	49	49	43	-3	3
15.	VD-C	24	38	20	21	31	28	12	13
16.	VD-A	27	45	30	29	35	25	16	5
17.	VD-DF	20	22	22	22	17	20	-5	3
18.	RR-D	24	36	32	43	35	35	7	3
19.	SF-Z	29	42	30	48	41	35	15	-11
20.	LD	48	54	43	35	38	38	10	5
21.	MS-D	27	18	16	23	10	10	-17	-17
22.	MS-LL	30	17	17	27	18	19	-12	-9
23.	MS-HL	24	17	18	26	22	17	-16	-9
24.	IS	25	26	24	17	32	26	-2	0
25.	SA-FR	43	44	30	35	28	30	-3	5
26.	SA-Z	7	6	19	17	12	11	-12	-11
27.	SA-D	20	21	24	26	26	27	16	14
28.	SA-O	16	19	11	17	9	11	7	12
29.	BF	7	8	1	-8	2	0	4	8
30.	V	17	17	9	20	13	16	-10	-1
31.	S	11	7	5	20	9	9	-7	1
32.	SAT-V	19	11	13	20	14	13	-3	5
33.	SAT-M	2	11	13	8	13	6	1	0

Table 3 - Page 2

Intercorrelations Among the 33 Variables for 200 Subjects

(decimals omitted)

		9 PA-C	10 PA-M	11 PA-11	12 PA-CA	13 SL-C	14 SL-M	15 VD-C	16 VD-A
1.	FR-C	53	42	48	44	41	48	24	27
2.	FR-S	53	47	46	41	51	46	38	45
3.	FR-CO	45	43	47	44	41	52	20	30
4.	FR-AB	44	42	45	47	43	49	21	29
5.	FR-11	48	43	50	51	40	49	31	35
6.	FR-CA	40	47	46	44	39	43	28	25
7.	CL-11	13	7	9	11	-9	-3	12	16
8.	CL-CA	9	18	13	13	-4	3	13	5
9.	PA-C	--	63	66	59	44	50	27	26
10.	PA-M	63	--	65	60	46	49	34	30
11.	PA-11	66	65	--	71	50	58	28	25
12.	PA-CA	59	60	71	--	44	52	21	28
13.	SL-C	44	46	50	44	--	59	29	18
14.	SL-M	50	49	58	52	59	--	35	25
15.	VD-C	27	34	28	21	29	35	--	55
16.	VD-A	26	30	25	28	18	25	55	--
17.	VD-DF	15	19	15	19	22	21	9	15
18.	RR-D	33	32	29	35	33	39	28	36
19.	SF-Z	26	31	27	37	34	38	34	36
20.	LD	37	34	36	30	29	28	35	38
21.	MS-D	4	9	16	16	23	12	0	11
22.	MS-I	7	15	19	22	30	20	10	8
23.	MS-HL	16	20	20	26	34	24	5	3
24.	IS	24	28	24	21	34	37	39	34
25.	SA-FR	26	25	32	32	25	22	17	13
26.	SA-Z	5	12	9	13	9	15	11	10
27.	SA-D	19	23	24	30	23	34	22	9
28.	SA-O	17	16	9	6	7	18	12	9
29.	BF	-7	-3	-1	0	1	-6	10	6
30.	V	22	18	24	31	31	18	4	2
31.	S	15	15	20	19	27	16	5	8
32.	SAT-V	21	18	17	31	14	16	-5	-1
33.	SAT-M	10	17	7	9	19	24	16	5

Table 3 - Page 3

Intercorrelations Among the 33 Variables for 200 Subjects

(decimals omitted)

		17 VD-DF	18 RR-D	19 SF-Z	20 LD	21 MS-D	22 MS-LL	23 MS-HL	24 IS
1.	FR-C	20	24	29	48	27	30	24	25
2.	FR-S	22	36	42	54	18	17	17	26
3.	FR-CO	22	32	30	43	16	17	18	24
4.	FR-AB	22	43	48	35	23	27	26	17
5.	FR-II	17	35	41	38	10	18	22	32
6.	FR-CA	20	35	35	38	10	19	17	26
7.	CL-II	-5	7	15	10	-17	-12	-16	-2
8.	CL-CA	3	3	-11	5	-17	-9	-9	0
9.	PA-C	15	33	26	37	4	7	16	24
10.	PA-M	19	32	31	34	9	15	20	28
11.	PA-II	15	29	27	36	16	19	20	24
12.	PA-CA	19	35	37	30	16	22	26	21
13.	SL-C	22	33	34	29	23	30	34	34
14.	SL-M	21	39	38	28	12	20	24	37
15.	VD-C	9	28	34	35	0	10	5	39
16.	VD-A	15	36	35	38	11	8	3	34
17.	VD-DF	--	11	9	17	15	25	27	16
18.	RR-D	11	--	55	30	16	10	9	16
19.	SF-Z	9	55	--	26	12	20	14	26
20.	LD	17	30	26	--	11	10	3	23
21.	MS-D	15	16	12	11	--	71	53	20
22.	MS-LL	25	10	20	10	71	--	63	27
23.	MS-HL	27	9	14	3	53	63	--	32
24.	IS	16	16	26	23	20	27	32	--
25.	SA-FR	1	12	25	36	13	16	12	0
26.	SA-Z	11	3	12	6	0	1	6	6
27.	SA-D	20	33	41	13	6	16	13	10
28.	SA-O	-2	8	13	3	6	4	0	10
29.	BF	1	-4	2	0	1	7	-6	1
30.	V	7	12	14	5	19	22	23	8
31.	S	3	21	13	0	27	31	35	11
32.	SAT-V	24	9	7	10	11	20	23	4
33.	SAT-M	27	6	5	10	13	24	27	36

Table 3 - Page 4

Intercorrelations Among the 33 Variables for 200 Subjects

(decimals omitted)

		25 SA-FR	26 SA-Z	27 SA-D	28 SA-O	29 BF	30 V	31 S	32 SAT-V	33 SAT-M
1.	FR-C	43	7	20	16	7	17	11	19	2
2.	FR-S	44	6	21	19	8	17	7	11	11
3.	FR-CO	30	19	24	11	1	9	5	13	13
4.	FR-AB	35	17	26	17	-8	20	20	20	8
5.	FR-II	28	12	26	9	2	13	9	14	13
6.	FR-CA	30	11	27	11	0	16	9	13	6
7.	CL-II	-3	-12	16	7	4	-10	-7	-3	1
8.	CL-CA	5	-11	14	12	8	-1	1	5	0
9.	PA-C	26	5	19	17	-7	22	15	21	10
10.	PA-M	25	12	23	16	-3	18	15	18	17
11.	PA-II	32	9	24	9	1	24	20	17	7
12.	PA-CA	32	13	30	6	0	31	19	31	9
13.	SL-C	25	9	23	7	1	31	27	14	19
14.	SL-M	22	15	34	18	-6	18	16	16	24
15.	VD-C	17	11	22	12	10	4	5	-5	16
16.	VD-A	13	10	9	9	6	2	8	-1	5
17.	VD-DF	1	11	20	-2	1	7	3	24	27
18.	RR-D	12	3	33	8	-4	12	21	9	6
19.	SF-Z	25	12	41	13	2	14	13	7	5
20.	LD	36	6	13	3	0	5	0	10	10
21.	MS-D	13	0	6	6	1	19	27	11	13
22.	MS-LL	16	1	16	4	7	22	31	20	24
23.	MS-HL	12	6	13	0	-6	23	35	23	27
24.	IS	0	6	10	10	1	8	11	4	36
25.	SA-FR	--	10	22	13	1	29	19	6	-22
26.	SA-Z	10	--	9	9	6	-6	-6	-3	-3
27.	SA-D	22	9	--	15	1	18	2	10	11
28.	SA-O	13	9	15	--	10	14	15	3	7
29.	BF	1	6	1	10	--	-2	17	-5	8
30.	V	29	-6	18	14	-2	--	55	50	9
31.	S	19	-6	2	15	-17	55	--	38	8
32.	SAT-V	6	-3	10	3	-5	50	38	--	31
33.	SAT-M	-22	-3	11	7	8	9	8	31	--

in keeping with the oft-made generalization that abilities are normally related, the number of positive correlations in Table 3 far exceed the number of negative correlations. Second, if we ask about correlations which are statistically different from zero (r = .18 at .01 level), we can see that the table has many which are this high and higher. At the very least, therefore, we have something to work with, although correlations as low as .18 will not be given much attention. At this point our intent is to identify variables which fail to correlate with other variables, and which may therefore be eliminated from the factor analysis.

Variables 7 and 8, which are the clustering measures for two of the free recall lists (FR-II and FR-CA), are not associated with other variables. The two clustering measures are strongly related, and they correlate at a low level with the learning scores from which they were derived, but beyond this about half the correlations involving these two variables are negative, and about half are positive.

Proceeding along the columns, we see that variable 17 (VD-DF), the double function verbal-discrimination list, failed to show a correlation higher than .27. Of particular interest is the fact that performance on this task was unrelated to the performances on the other two verbal-discrimination tasks.

Variable 24 (IS), a task intended to measure the suscepti-

bility to interference, does not show a strong relationship with any of the other tasks, although it shows weak relationships with many. We retained this variable in some of our preliminary factor analyses and it always failed to load in a decisive manner on any factor, so we have not included it in our final factor analysis.

Variables 25 through 28 represent the four tasks which were learned simultaneously. The free recall task (SA-FR), and the recognition task (SA-D) show some meaningful correlations, but this is not true for the other two tasks. For Task SA-Z, the subjects made frequency judgments of state names but the scores do not correlate with any other task, including the other frequency-judging task (SF-Z, Task 19). The ordering task (SA-O) also failed to relate to the scores on the other tasks. This may have been due to low reliability.

Variables 29 through 33 were identified as measures of semantic memory. They will not be included in our factor analysis of episodic memory tasks but the correlations between the two types of tasks should be examined. One fact which stands out in such an examination is that our episodic memory tasks and the semantic memory tasks represent two different worlds. Scores on the vocabulary test (Task 30) have some suggestive relationships with scores from some of the episodic memory tasks (e.g., PA-CA; SL-C), but they are most strongly related to other semantic memory tasks (spelling and SAT verbal). There are some scattered but low relationships between the other tests (31-33)

and the episodic tests, but clearly there are no systematic trends indicated by the correlations.

Factor Structure

After having eliminated variables 7, 8, 17, 24, 26 and 28, we had 22 variables remaining to represent performance on episodic memory tasks. To characterize the meaning of the remaining measures, various factor analytic techniques were used. The basic strategy was first to estimate number of factors using principal components, then to use a maximum likelihood approach (Jöreskog, 1967) to estimate factor loadings. The latter assumes that a linear model underlies the data, and yields a solution with well defined properties and a large sample goodness-of-fit test. In addition, a cluster analytic and principle components approaches were taken to establish whether notable differences among results appeared. The criteria for choice for the final solution included consistency of results across approaches, and formal standards for judging the quality of the solution, such as goodness-of-fit tests.

The preliminary principle components analysis yielded five eigenvalues whose magnitudes were greater than 1.00, namely 8.41, 2.13, 1.46, 1.30, 1.14. The maximum likelihood for estimates of factor loadings for a five factor model are given in Table 4. It will be remembered that the correlations given in Table 3 were notable

Table 4

Rotated Orthogonal Factor Matrix Resulting from the Maximum Likelihood Technique.

The Scores on the 22 Tasks were Based on 200 Subjects.

	Task	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
1.	FR-C	30	69	22	5	17
2.	FR-S	24	70	8	14	40
3.	FR-CO	26	74	7	22	5
4.	FR-AB	24	61	17	43	3
5.	FR-II	32	60	6	35	10
6.	FR-CA	29	55	7	32	9
25.	SA-FR	18	39	11	6	15
20.	LD	19	47	2	3	42
9.	PA-C	66	38	-2	9	18
10.	PA-M	67	27	6	17	23
11.	PA-II	80	29	10	13	11
12.	PA-CA	70	25	14	28	7
13.	SL-C	43	32	24	24	15
14.	SL-M	50	35	11	35	12
15.	VD-C	18	10	1	23	63
16.	VD-A	10	21	1	22	63
18.	RR-D	17	19	3	58	26
27.	SA-D	18	10	9	46	5
19.	SF-Z	11	20	10	67	28
21.	MS-D	1	11	77	2	6
22.	MS-LL	6	9	90	10	4
23.	MS-HL	16	11	67	10	-6

for the rather strong relationships which existed among the scores on the free-recall, paired-associate, and serial tasks. In spite of this, two factors emerged from the analysis. Factor 1 in Table 4 is associated with the paired-associate tasks and, to a lesser extent, to the two serial tasks. Factor 2, on the other hand, includes all the free-recall tasks plus the task used to measure list discrimination (Task 20, LD). Factor 3 is particularly clean in that the loadings for tasks other than the memory-span tasks were low. Factor 4, although most clearly identified with recognition and frequency discrimination, did not escape loadings of some magnitude from many of the other tasks. A similar description can be given for Factor 5 in that the strongest component came from the verbal-discrimination tasks but many of the other tasks loaded to lesser degrees. A first general conclusion is that the factors appear to be strongly associated with the type of tasks.

The chi square test showed a highly significant discrepancy between the five factor model and the data ($p = .001$). This could be due to the fact that the amount of variance encompassed by the five factors was appreciably less than the total variance. If this is true, forcing additional factors should reduce the discrepancy between the model and the data. When six factors were extracted, the chi square remained significant ($p = .004$), but when seven factors were extracted, the chi square value would be judged not to represent a significant discrepancy ($p = .08$). The chi square test is <u>not</u> the

only criterion which ought to be used in forming opinions about the quality of a maximum likelihood solution. The next criteria include the meaningfulness of the solutions and size of residuals. Consider first the meaningfulness of the solution.

The five factors as extracted are somewhat disappointing conceptually in that none of the attributes we attempted to measure indirectly produced unique individual differences of such magnitude as to represent a factor. For example, in spite of the fact that the effect of concrete words was large in free-recall learning, there was no indication of this in the factor structure. Indeed, when only the seven free-recall tasks were submitted to a factor analysis, only one factor emerged. When we forced seven factors from the 22 variables, we found that the serial tasks formed a factor, and the other factor consisted of the two free-recall tasks involving direct associations between words (Task 5, FR-II) and conceptual associations (Task, 6, FR-CA). This could be thought of an an implicit associational response (IAR) factor. Still, the question remained as to the reliability of the sixth and seventh factors. We then used the two methods involving principal component solutions, forcing seven factors. The IAR factor failed to appear in either analysis; the serial factor appeared in one. We then drew three random samples of 100 subjects from among the 200 and carried out a maximum likelihood analysis with seven factors specified. The IAR factor did not appear in any of the three samples, and no other attribute produced a factor. In fact, some of the so-called factors appeared to be

without meaning. All of this points toward the conclusion that the five-factor solution as given in Table 4 is the most appropriate one for our data.

Consider next the residuals. We examined the differences between the predicted correlations among the 22 tasks based on the five-factor solution and those actually obtained. These residuals were for the most part very small. Of the 231 residual correlations, only 13.9% were greater than .10. Of these, only two had large magnitudes. These two involved the correlations between Tasks 5 and 6 (FR-II and FR-CA) and those between Tasks 11 and 12 (PA-II and PA-CA). The residual correlation for Tasks 5 and 6 was .38, that for Tasks 11 and 12, .26. All four of these tasks involve IARs, but we have been unable to give this fact psychological meaning. It is likely, however, that the significant chi square noted earlier was to a large extent determined by the discrepancy between the predicted and obtained scores for these tasks.

We examined the results for two other techniques using principal component solutions. The first principal component solution involved having a 1.00 in each cell of the diagonal of the correlation matrix, and the second solution used multiple correlations in these cells as a means of estimating communalities. The correspondence in the outcomes for the three techniques (maximum likelihood and two using principal component solutions) was high; the number of factors which evolved was the same (five), and the tasks making up each factor were the same. Furthermore, the loadings of the individual tasks on the factors were quite similar among the three methods.

An oblique solution did not change anything fundamentally.

We carried out a cluster analysis (Revelle, in press), an approach which does not assume an underlying linear model, and found the outcome to be quite compatible with the five-factor solution. We applied the criterion developed by Montanelli and Humphreys (1976) and found that according to their methods five was an appropriate number of factors to extract. Roughly speaking, the criterion involves comparing the actual correlation matrix to one generated randomly in order to specify the number of factors.

In conclusion, we believe that the five-factor solution is an appropriate one for these 22 episodic memory tasks.

Discussion

Previous Work

It is necessary initially to make some comments about the degree to which our results correspond to those produced by previous investigators. There are only a few studies available in which the classes of tasks used overlap appreciably with those of the present study. Three sets of data may be used to characterize previous work. Anastasi (1930; 1932) used paired associates, free recall, recognition, and memory span. Kelley (1964) used paired associates, recognition, and memory span, while Botwinick and Storandt (1974) used paired associates, serial lists, recognition, and memory span. In all studies the measures of memory span stood isolated from the measures on the other tasks, just as was true in our study. Less clear was the independence between associative

tasks (paired associates; serial; free recall) and recognition tasks. Nevertheless, the results of our study are quite representative. A correlation matrix will normally show some statistically significant correlations between associative learning and recognition, but if the matrix is subjected to a factor analysis, the associative tasks and the recognition tasks form different factors. And, as was true in Table 3, the associative tasks correlate much higher among themselves than they correlate with recognition tasks.

There were so many differences in method or procedure among the above studies and between them and the present study that the relatively high correspondence in the findings is impressive. Some of these differences in method may be noted. The ages of the subjects used by Botwinick and Storandt varied between 21 and 80. Kelley's subjects were Air Force Cadets, and all of the tests were given on a single day. Some of the tasks included nonsense syllables, some included words. Thus, looking across investigations it seems that the particular type of task is a powerful determinant of individual differences, and that variations within a task type are relatively impotent. Except for two cases (to be discussed later), our data show this.

Our data showed that there was no relationship between semantic memory as measured by a vocabulary test and performance on the episodic memory tests. Two of the above investigators (Anastasi; Kelley) also gave their subjects vocabulary tests and they too failed to find a relationship between the two sets of measurements. We will

conclude that our results are in essential correspondence with the work of these previous investigators.

Attribute Measurement

One of the purposes of our study was to determine if attributes of memory could be measured as independent factors. The frequency attribute, of course, was measured directly. Our concern at this point is with the attributes that we attempted to measure indirectly. We were unable to demonstrate unique individual differences on the acoustic attribute, on the imagery attribute, on the implicit associative attribute, and so on. In some cases our independent variable failed to influence overall rate of learning, thus denying the possibility of finding individual differences on the attribute assumed to be associated with the independent variable. In other cases, the independent variable had the expected influence but still the analysis did not show that the attribute produced (or was associated with) a unique source of variance. Interitem associations produced a huge effect on free-recall learning but there was no evidence that the relative magnitude of the effect differed greatly among individuals. To say this another way, it appears that the fundamental basis or source of individual differences was not changed by the presence of strong interitem associations in the free-recall lists. Indeed, because we failed to find unique variance associated with the manipulated variable which produced large effects on learning, we may be led to believe that the same conclusion would have been reached had

all of our independent variables had their expected influence on learning. These findings are pointing toward two conclusions which we will now discuss.

Subject control of attributes. In the running-recognition task we sought to measure individual differences in the susceptibility to false alarms on experimentally controlled words. It was a surprise to discover that on the second list the number of false alarms on these words was actually less than the number on control words. It was also found that crossed associates and conceptual pairings in paired-associate lists (PA-II and PA-CA) produced little if any interference in learning in spite of the fact that we knew from other studies that rather heavy interference could be anticipated, at least with naive subjects. These findings, when viewed in conjunction with a number of previous studies (Zimmerman, Shaughnessy, & Underwood, 1972; Galbraith, 1975; Underwood, Reichardt, and Malmi, 1975), seem to lead to the conclusion that subjects have the capacity to select attributes that are appropriate to the demands of the task, and to set aside those attributes that are not appropriate. We presume that a part of what is meant by being well practiced is the skill involved in selecting the most appropriate attributes for performing a given task. For example, we observed a very large positive effect of interitem associations and of conceptual associations in free-recall learning. In this task, if the subject "allows" the associative attributes to dominate performance, performance is facilitated. We will later point out another case in our data where

it appears that subjects were able to control the particular attribute determining their responses. For now we conclude that the capacity of subjects to control attribute selection reminds us again that we must distinguish between attributes present in memory and attributes which determine memory performance.

Dominance of associative processes. Another matter which would work against the detection of individual differences in attribute functioning can be identified in paired-associate, serial, and free-recall learning. The underlying individual differences in rate of associative learning appear to be so powerful that they dominate and obscure any relatively small amounts of variance due to individual differences on another factor, even if such exists. This should be illustrated at the risk of repeating a line of thought developed much earlier. Concrete words resulted in much more rapid learning than did abstract words. We relate this difference to the imagery attribute. The problem lies in the correlation of the learning scores for the concrete list and the scores for the abstract list, in our study the value being .65, which is essentially as high as the reliability of either task. This correlation means either of two things. First, it may mean that there are individual differences in the influence of the imagery attribute, but overall learning rate may be (for example) 90% determined by associative learning which is independent of the imagery attribute, and only 10% due directly to the imagery attribute. Thus, individual differences due to associative learning dom-

inate the variance in the data. The second possibility is that imagery differences among individuals are highly correlated with individual differences in associative learning. This is to say that imagery is not a unique source of individual differences. We do not believe that our data allow a choice between the two, although we lean toward the first possibility as being the most likely reason for the failure of imagery to emerge as a reliable attribute.

Theoretical Implications

One of the reasons for undertaking the present study was as a means of providing evidence for disciplining and sharpening theoretical thinking when the tasks are used in experimental research. We believe that some of our results are particularly pointed with regard to this objective.

At a general level there is reasonably good evidence to support a distinction between attributes which discriminate among memories and those that retrieve memories. This is an age-old issue and it may be that some will judge our data to be indefinite on the issue because there were in fact statistically significant correlations between performance on retrieval tasks and performance on the recognition tasks, particularly Task 18, RR-D. Thus it could be said that the recognition factor (Factor 4) is by no means completely independent of so-called associative tasks such as free recall. Yet, it would seem quite inappropriate to ignore the differences in the magnitude of correlations, hence to ignore the fact that associative tasks and recognition tasks form two different factors. Somehow this must

be reflected in theory.

Within the discriminative attributes, some of the findings appear to be very troublesome for extant theory, particularly frequency theory. This theory was originated to account for verbal-discrimination learning (Ekstrand, Wallace, & Underwood, 1966), and was subsequently extended to classical recognition memory (Underwood, 1971). Now, we find that the scores on two verbal-discrimination tasks form one factor, while the scores on the recognition tasks form another. Perhaps never before has a theory which evolved from experimental work been so savagely attacked by a correlational approach. However, the fact that scores on the frequency-discrimination task (SF-Z, Task 19) most heavily load on the factor with recognition tasks indicates that the problem for the theory lies in the verbal-discrimination task. There may be a solution to the problem along the lines to follow.

Ghatala, Levin, and Subkoviak (1975) have shown that by a rather simple instructional procedure subjects may abandon the frequency attribute in verbal-discrimination learning and base their performance on a different strategy. More particularly, the subject learns a two-category classification in which incorrect words are placed in one category, correct words in another. In our case, many of our subjects may have classified items as those underlined and those not underlined. Ghatala et. al. (1975) demonstrated that verbal-discrimination learning was as rapid, or even more rapid (with some groups), when learning occurred by the two-category method than by frequency discrimination between the two words in a pair. We have

no evidence that our subjects did turn the verbal-discrimination task into a two-category learning task, but it is not unreasonable to suppose that this happened.

There is a related issue involved in the case of the temporal attribute as exemplified by the LD task (Task 20). Table 4 shows that the scores on this task load both on the free-recall factor and on the verbal-discrimination factor. We have presumed that category learning underlies temporal differentiation as exemplified in the LD Task (Underwood, 1977). In our LD task, according to this point of view, a subject established three categories, one for each list. It seems reasonable, therefore, that the scores on the LD and VD tasks should be correlated. This cannot be the whole story because the LD task also loads on the free-recall factor, whereas the VD task does not.

We have already noted that the memory-span tasks did not relate strongly to other tasks, and that this same finding has been reported by others. Indeed, a recent study by Hunt, Lunneborg, and Lewis (1975) has shown that the digit span is not related to any one of a great variety of tasks, including tasks that are commonly said to be techniques for measuring short-term memory. In our study, two other episodic tasks had the same status as the memory-span tests in that they did not correlate strongly with the memory-span tests nor with tasks falling in the other four factors. The two tasks in question were the double function, verbal-discrimination task (VD-DF), Task 17), and the task constructed to measure susceptibility to interference between associations established within the laboratory

(IS, Task 24). We will examine each of these in order.

Task 17, VD-DF, is a verbal-discrimination task in mode of presentation only, since it does not correlate with the other verbal-discrimination tasks. Furthermore, associative learning, as represented by Factors 1 and 2, are at best only weakly involved. When this task was included in an early factor analysis it loaded most heavily (.35) on a factor that was most clearly identified (.69) by the mathematics part of the SAT. Given such a relationship, it is almost always possible to devise a reason why the relationship should exist, but it is perhaps better to dismiss the whole matter and simply indicate that we do not know what processes underly the learning which occurs for the double-function list.

The other unique task, the IS task, is a paired-associate task in form only because it does not correlate very highly with the other paired-associate tasks (see Table 3). When this task was included in a factor analysis, it loaded about equally on the verbal-discrimination factor, and on the factor dominated by the mathematics test. These loadings were not high and we have tended to disregard them. Our own belief is that this task measures what it was constructed to measure, susceptibility to interference. We had expected that tasks PA-II and PA-CA (Tasks 11 and 12) would also show interference. Because they did not, we wound up without another task that could be expected to be related to the IS task. The group data for this task indicated that interference did build up across lists within sets (Figure 3), and we are inclined toward the position that the indivi-

dual differences on the task do represent in part the individual differences in susceptibility to interference.

The free-recall tasks and the paired-associate tasks formed two factors, and the serial task fell within the paired-associate factor. Nevertheless, the rather substantial correlations among these three tasks (see Table 3) indicate an underlying commonality. We have assumed that the common component is the need in all tasks to establish associations and that the process of establishing associations is independent of tasks. Still, two factors emerged and in the long run it will be necessary to try to understand why this should be so. There are obvious differences between free-recall and paired-associate tasks, but we have no insights that tell us how these differences would lead to a separate factor. We sometimes have wondered if we should think about different kinds of associations. Assume that contextual associations are involved in free recall. Are these associations in some way different from the associations established between two words in a paired-associate list? We believe that two-category classification learning requires associative learning, but are the associations to be viewed in the same way we view associations in other tasks?

One final matter should be mentioned in this section. It has been noted that our measures of semantic memory and the SAT scores showed weak relationships at best with the performance on the episodic memory tasks. It might be expected that vocabulary skills would be rather sharply related to performances on episodic memory tasks using

words. This was obviously not the case. It must be understood that positive correlations need not be expected because the scores on the vocabulary tests are a function of two variables, namely, ability to learn and time spent in learning. In the episodic memory tasks, time is constant and only ability is allowed to produce individual differences. The critical point is that it should not be concluded that the skills or abilities measured by episodic tasks have nothing to do with the skills leading to the information available in semantic memory. Suppose, for example, that we had included among our tasks a paired-associate list in which the stimulus words were very rare words (the meaning of which the subjects would not know) and the response terms were synonyms of the stimulus terms. It seems beyond doubt that the performance on this task would be highly correlated with the paired-associate tasks we actually used.

Overall Perspective

Looking at our results in a most general way, we offer two conclusions concerning the manner in which the data have changed our perspective. These conclusions depend initially upon the division of tasks into recall-like tasks and recognition-like tasks. The attribute conception of memory stresses the multi-attribute nature of memories. Within this conception the theoretical problem lies in specifying the role played by each attribute in memory functioning. Let us see how this conception may be viewed now when applied to the recognition-like tasks (attributes which discriminate among memories).

The evidence we have obtained, as well as the evidence which is being generated by other investigators, begins to question this approach. How can we have a simple and strightforward theory about the functions of a particular attribute when a subject, almost capriciously, decides to change the attribute that dominates his performance? How can we have a straightforward theory about how the frequency attribute mediates recognition memory and verbal-discrimination learning when a relatively minor change in instructions can cause a subject to ignore frequency information and transfer his attention to another attribute; or to a form of associative learning?

The second conclusion places the focus on associative learning. If we now see that some of the attributes are unstable for theoretical purposes, we can at the same time see the contrast in the stability of the associative learning processes. Tasks which require formation of associations do not allow the subject many ways to be capricious, and differences in the rate at which associations are formed represent a fundamental individual difference variable.

References

Anastasi, A. A group factor in immediate memory. Archives of Psychology, 1930, 18, No. 120.

Anastasi, A. Further studies on the memory factor. Archives of Psychology, 1932, 22, No. 142.

Battig, W. F., & Montague, W. E. Category norms for verbal items in 56 categories: A replication and extension of the Connecticut category norms. Journal of Experimental Psychology Monograph, 1969, 80 (3, Pt. 2).

Biers, D. W. The effect of frequency, affect, and total time in verbal-discrimination learning. Unpublished doctoral dissertation, Northwestern University, 1970.

Botwinick, J., & Storandt, M. Memory, related functions and age. Springfield, Illinois: C. C. Thomas, 1974.

Chiang, A., & Atkinson, R. C. Individual differences and interrelationships among a select set of cognitive skills. Memory and Cognition, 1976, 4, 661-672.

Conrad, R. Acoustic confusions in immediate memory. British Journal of Psychology, 1964, 55, 75-84.

Cronbach, L. J., & Furby, L. How should we measure "change" -- or should we? Psychological Bulletin, 1970, 74, 69-80.

Ekstrand, B. R., Wallace, W. P., & Underwood, B. J. A frequency theory of verbal-discrimination learning. Psychological Review, 1966, 73, 566-578.

Galbraith, R. C. On the independence of attributes of memory. Journal of Experimental Psychology: Human Learning and Memory. 1975, 1, 23-30.

Ghatala, E. S., Levin, J. R., & Subkoviak, M. J. Rehearsal strategy effects in children's discrimination learning: Confronting the crucible. Journal of Verbal Learning and Verbal Behavior, 1975, 14, 398-407.

Heise, D. R. Semantic differential profiles for 1,000 most frequent English words. Psychological Monographs, 1965, 79, (8, Whole No. 601).

Hunt, E., Lunneborg, C., & Lewis, J. What does it mean to be high verbal? Cognitive Psychology, 1975, 7, 194-227.

Jöreskog, K. G. Some contributions to maximum likelihood factor analysis. Psychometrika, 1968, 33, 37-72.

Kelley, H. P. Memory abilities: A factor analysis. Psychometric Monographs, 1964, No. 11.

Paivio, A., Yuille, J. C., & Madigan, S. A. Concreteness, imagery, and meaningfulness values for 925 nouns. Journal of Experimental Psychology Monograph, 1968, 76, (1, Pt. 2).

Revelle, W. Hierarchical cluster analysis and the internal structure of tests. Multivariate Behavioral Research, in press.

Shepard, R. N., & Teghtsoonian, M. Retention of information under conditions approaching a steady state. Journal of Experimental Psychology, 1961, 62, 302-309.

Thorndike, E. L. & Lorge, I. The teacher's word book of 30,000 words. New York: Teachers College, Bureau of Publications.

Tulving, E. Episodic and semantic memory. In E. Tulving & W. Donaldson (Eds.), Organization of memory. New York: Academic Press, 1972.

Underwood, B. J. Attributes of memory. Psychological Review, 1969, 76, 559-573.

Underwood, B. J. Recognition memory. In H. H. Kendler and J. T. Spence (Eds.), Essays in neobehaviorism. New York: Appleton-Century-Crofts, 1971.

Underwood, B. J. Temporal codes for memories: Issues and problems. Hillsdale, New Jersey: L. Erlbaum Associates, 1977.

Underwood, B. J., & Ekstrand, B. R. Linguistic associations and retention. Journal of Verbal Learning and Verbal Behavior, 1968, 7, 162-171.

Underwood, B. J., & Reichardt, C. S. Contingent associations and the double-function, verbal-discrimination task. Memory & Cognition, 1975, 3, 311-314.

Underwood, B. J., Reichardt, C. S., & Malmi, R. A. Sources of facilitation in learning conceptually structured paired-associate lists. Journal of Experimental Psychology: Human Learning and Memory, 1975, 1, 160-166.

Underwood, B. J., & Schulz, R. W. Studies of distributed practice: XXI. Effect of interference from language habits. Journal of Experimental Psychology, 1961, 62, 571-575.

Zimmerman, J., Broder, P. K., Shaughnessy, J. J., / Underwood, B. J. A recognition test of vocabulary using signal-detection measures and some correlates of word and nonword recognition. Intelligence, 1, 1977, in press.

Zimmerman, J., Shaughnessy, J. J., & Underwood, B. J. The role of associations in verbal-discrimination learning. American Journal of Psychology, 1972, 85, 499-518.

Footnote

We wish to thank Susan Kapelak for testing all of the subjects, and to thank Dr. William R. Revelle for his aid in carrying out a cluster analysis of the data.

DISTRIBUTION LIST

Navy

4 Dr. Marshall J. Farr, Director
Personnel & Training Research Programs
Office of Naval Research (Code 458)
Arlington, VA 22217

1 ONR Branch Office
495 Summer St.
Boston, MA 02210
Attn: Dr. James Lester

1 ONR Branch Office
1030 E. Green St.
Pasadena, CA 91101
Atten: Dr. Eugene Gloye

1 ONR Branch Office
536 S. Clark St.
Chicago, IL 60605
Atten: Dr. C. E. Davis

1 Dr. M. A. Bertin, Scientific Director
Office of Naval Research
Scientific Liaison Group/Tokyo
American Embassy
APO San Francisco 96503

1 Office of Naval Research
Code 200
Arlington, VA 22217

6 Commanding Officer
Naval Research Laboratory
Code 2627
Washington, DC 20390

1 LCDR Charles J. Theisen, Jr., MSC, USN
4024
Naval Air Development Center
Warminster, PA 18974

1 Commanding Officer
U.S. Naval Amphibious School
Coronado, CA 92155

1 Commanding Officer
Naval Health Research Center
San Diego, CA 92152
Attn: Library

1 Chairman, Leadership & Law Dept.
Div. of Professional Development
U. S. Naval Academy
Annapolis, MD 21402

1 Scientific Advisor to the Chief of
Naval Personnel (Pers Or)
Naval Bureau of Personnel
Room 4410, Arlington Annex
Washington, DC 20370

1 Dr. J. R. Borsting
Provost & Academic Dean
U.S. Naval Postgraduate School
Monterey, CA 93940

1 Mr. Maurice Callahan
NODAC (Code 2)
Dept. of the Navy
Bldg. 2, Washington Navy Yard
(Anacostia)
Washington, DC 20374

1 Superintendent (Code 1424)
Naval Postgraduate School
Monterey, CA 93940

1 Chief of Naval Technical Training
Naval Air Station Memphis (75)
Millington, TN 38054
Attn: Dr. Norman J. Kerr

1 Principal Civilian Advisor for
Education & Training
Naval Training Command, Code 00A
Pensacola, FL 32508
Attn: Dr. W. L. Maloy

1 Dr. A. F. Smode, Director
Training Analysis & Evaluation Group
Dept. of the Navy
Orlando, FL 32813

1 Chief of Naval Education & Training
Support (01A)
Pensacola, FL 32509

1 Capt. H. J. Connery, USN
Navy Medical R&D Command
NNMC, Bethesda, MD 20014

1 Navy Personnel R&D Center
Code 01
San Diego, CA 92152

2 Navy Personnel R&D Center
Code 306
San Diego, CA 92152
Attn: Dr. James McGrath

5 A. A. Sjoholm, Head, Technical Support
Navy Personnel R & D Center
Code 201
San Diego, CA 92152

1 Navy Personnel R&D Center
San Diego, CA 92152
Attn: Library

1 Navy Personnel R&D Center
San Diego, CA 92152
Attn: Dr. J. D. Fletcher

1 Capt. D. M. Gragg, MC, USN
Head, Section on Medical Education
Uniformed Services Univ. of the
Health Sciences
6917 Arlington Rd.
Bethesda, MD 20014

1 LCDR J. W. Snyder, Jr.
F-14 Training Model Manager
VF-124
San Diego, CA 92025

1 Dr. John Ford
Navy Personnel R&D Center
San Diego, CA 92152

1 Dr. Worth Scanland
Chief of Naval Education & Training
NAS, Pensacola, FL 32508

Army

1 Technical Director
U.S. Army Research Institute for the
Behavioral & Social Sciences
1300 Wilson Blvd.
Arlington, VA 22209

1 Armed Forces Staff College
Norfolk, VA 23511
Attn: Library

1 Commandant
U.S. Army Infantry School
Ft. Benning, GA 31905
Attn: ATSH-I-V-IT

1 Commandant
U.S. Army Institute of Administration
Attn: EA
Ft. Benjamin Harrison, IN 46216

1 Dr. F. J. Harris
U.S. Army Research Institute
1300 Wilson Blvd.
Arlington, VA 22209

1 Dr. Ralph Dusek
U.S. Army Research Institute
1300 Wilson Blvd.
Arlington, VA 22209

1 Dr. Leon Nawrocki
U.S. Army Research Institute
1300 Wilson Blvd.
Arlington, VA 22209

1 Dr. Joseph Ward
U.S. Army Research Institute
1300 Wilson Blvd.
Arlington, VA 22209

1 Dr. M. S. Katz, Chief
Individual Training & Performance
Evaluation Technical Area
U.S. Army Research Institute
1300 Wilson Blvd.
Arlington, VA 22209

1 HQ USAREUE & 7th Army
ODCSOPS
USAREUR Director of GED
APO New York 09403

1 ARI Field Unit - Leavenworth
P.O. Box 3122
Ft. Leavenworth, KS 66027

1 DCDR, USAADMINCEN
Bldg. #1, A310
Atten: AT21-OED Library
Ft. Benjamin Harrison, IN 46216

1 Dr. James Baker
U.S. Army Research Institute
1300 Wilson Blvd.
Arlington, VA 22209

Air Force

1 Research Branch
AFMPC/DPMYP
Randolph AFB, TX 78148

1 AFHRL/AS (Dr. G. A. Eckstrand)
Wright-Patterson AFB
Ohio 45433

1 Dr. R. L. Morgan (AFHRL/ASR)
Wright-Patterson AFB
Ohio 45433

1 Dr. Marty Rockway (AFHRL/TT)
Lowry AFB
Colorado 80230

1 Instructional Technology Branch
AFHRL
Lowry, AFB, CO 80230

1 Dr. A. R. Fregly
AFOSR/NL, Bldg. 410
Bolling AFB, DC 20332

1 Dr. Sylvia R. Mayer (MCIT)
HQ Electronic Systems Division
LG Hanscom Field
Bedford, MA 01730

1 Capt. Jack Thorpe, USAF
AFHRL/FTS
Williams AFB, AZ 85224

1 Air University Library
AUL/LSE 76-443
Maxwell AFB, AL 36112

Marine Corps

1 Director, Office of Manpower Utilization
HQ, Marine Corps (Code MPU)
BCB, Bldg. 2009
Quantico, VA 22134

1 Dr. A. L. Slafkosky
Scientific Advisor (Code RD-1)
HQ, U.S. Marine Corps
Washington, DC 20380

1 AC/S, Education Programs
Education Center, MCDEC
Quantico, VA 22134

Coast Guard

1 Mr. J. J. Cowan, Chief
Psychological Research Branch (G-P-1/62)
U.S. Coast Guard Headquarters
Washington, DC 20590

Other DOD

1 Dr. H. F. O'Neil, Jr.
Advanced Research Projects Agency
Cybernetics Technology, Rm. 623
1400 Wilson Blvd.
Arlington, VA 22209

1 Dr. Robert Young
Advanced Research Projects Agency
1400 Wilson Blvd.
Arlington, VA 22209

12 Defense Documentation Center
Cameron Station, Bldg. 5
Alexandria, VA 22314
Attn: TC

1 Military Assistant for Human Resources
Office of the Director of Defense Research & Engineering
Rm. 3D129, The Pentagon
Washington, DC 20301

1 Director, Management Information
Systems Office
OSD, M&RA
Rm. 3B917, The Pentagon
Washington, DC 20301

Other Government

1 Dr. Vern Urry
Personnel R&D Center
U.S. Civil Service Commission
1900 E St., NW
Washington, DC 20415

1 Dr. A. R. Molnar
Science Education Dev. & Res.
National Science Foundation
Washington, DC 20550

1 Dr. M. S. Smith, Assoc. Director
NIE/OPEPA
National Institute of Education
Washington, DC 20208

1 Dr. J. L. Young, Director
Memory & Cognitive Processes
National Science Foundation
Washington, DC 2055

Miscellaneous

1. Dr. J. R. Anderson
Dept. of Psychology
Yale University
New Haven, CT 06520

1 Dr. Scarvia B. Anderson
Educational Testing Service
Suite 1040
3445 Peachtree Rd., NE
Atlanta, GA

1 Dr. R. K. Branson
1A Tully Bldg.
Florida State University
Tallahassee, FL 32306

1 Dr. J. S. Brown
Bolt Beranek & Newman, Inc.
50 Moulton St.
Cambridge, MA 02138

1 Dr. R. P. Carver
School of Education
University of Missouri
5100 Rockhill Rd.
Kansas City, MO 64110

1 Century Research Corp.
4113 Lee Highway
Arlington, VA 22207

1 Dr. K. E. Clark
College of Arts & Sciences
University of Rochester
River Campus Station
Rochester, NY 14627

1 Dr. A. M. Collins
Bolt Beranek & Newman, Inc.
50 Moulton St.
Cambridge, MA 02138

1 Dr. J. J. Collins
Essex Corp.
6305 Caminito Estrellado
San Diego, CA 92120

1 Dr. Rene V. Dawis
Dept. of Psychology
University of Minnesota
Minneapolis, MN 55455

1 Dr. Ruth Day
Dept. of Psychology
Yale University
Box 11A, Yale Station
New Haven, CT 06520

1 ERIC Facility-Acquisitions
4833 Rugby Ave.
Bethesda, MD 20014

1 Major I. N. Evonic
Canadian Forces Personnel Applied
Research Unit
1107 Avenue Rd.
Toronto, Ontario, CANADA

1 Dr. Victor Fields
Dept. of Psychology
Montgomery College
Rockville, MD 20850

1 Dr. E. A. Fleishman
Advanced Research Resources Organization
8555 Sixteenth St.
Silver Spring, MD 20910

1 Dr. J. R. Frederiksen
Bolt Beranek & Newman, Inc.
50 Moulton St.
Cambridge, MA 02138

1 Dr. V. S. Gerlach
College of Education
146 Payne Bldg. B
Arizona State University
Tempe, AZ 85281

1 Dr. Robert Glaser, Co-Director
University of Pittsburgh
3939 O'Hara St.
Pittsburgh, PA 15213

1 Dr. M. D. Havron
Human Sciences Research, Inc.
7710 Old Spring House Rd.
West Gate Industrial Park
McLean, VA 22101

1 HumRRO/Western Division
27857 Berwick Dr.
Carmel, CA 93921
Attn: Library

1 HumRRO/Columbus Office
Suite 23,
2601 Cross Country Dr.
Columbus, GA 31906

1 HumRRO/Ft. Knox Office
P.O. Box 293
Fort Knox, KY 40121

1 HumRRO/Western Division
27857 Berwick Dr.
Carmel, CA 93921
Attn: Dr. Robert Vineberg

1 Dr. L. B. Johnson
Lawrence Johnson & Associates, Inc.
Suite 502
2001 S St., NW
Washington, DC 20009

1 Dr. Arnold F. Kanarick
Honeywell, Inc.
2600 Ridgeway Pkwy.
Minneapolis, MN 55413

1 Dr. R. A. Kaufman
203 Dodd Hall
Florida State University
Tallahassee, FL 32306

1 Dr. S. W. Keele
Dept. of Psychology
University of Oregon
Eugene, OR 97403

1 Dr. David Klahr
Dept. of Psychology
Carnegie-Mellon University
Pittsburgh, PA 15213

1 Dr. R. R. Mackie
Human Factors Research, Inc.
6780 Corton Dr.
Santa Barbara Research Park
Goleta, CA 93017

1 Dr. W. C. Mann
University of Southern California
Information Sciences Institute
4676 Admiralty Way
Marina Del Rey, CA 90291

1 Dr. Leo Munday
Houghton Mifflin Co.
P.O. Box 1970
Iowa City, IA 52240

1 Dr. D. A. Norman
Dept. of Psychology C-009
University of California, San Diego
La Jolla, CA 92093

1 Mr. A. J. Pesch, President
Eclectech Associates, Inc.
P. O. Box 178
N. Stonington, CT 06359

1 Mr. Luigi Petrullo
2431 N. Edgewood St.
Arlington, VA 22207

1 Dr. Steven M. Pine
N 660 Elliott Hall
University of Minnesota
75 East River Rd.
Minneapolis, MN 55455

1 R. Dir. M. Rauch
P II 4
Bundesministerium der Verteidigung
Postfach 161
53 Bonn 1, GERMANY

1 Dr. J. W. Rigney
University of Southern California
Behavioral Technology Laboratories
3717 South Grand
Los Angeles, CA 90007

1 Dr. A. M. Rose
American Institutes for Research
1055 Thomas Jefferson St., NW
Washington, DC 20007

1 Dr. L. L. Rosenbaum, Chairman
Dept. of Psychology
Montgomery College
Rockville, MD 20850

1 Dr. M. D. Reckase
Educational Psychology Dept.
University of Missouri-Columbia
12 Hill Hall
Columbia, MO 65201

1 Dr. R. J. Seidel
Instructional Technology Group, HumRRO
300 N. Washington St.
Alexandria, VA 22314

1 Dr. Richard Snow
Stanford University
School of Education
Stanford, CA 94305

1 Mr. D. J. Sullivan
c/o Canyon Research Group, Inc.
32107 Lindero Canyon Rd.
Westlake Village, CA 91360

1 Dr. C. R. Vest
Battelle Memorial Institute
Washington Operations
2030 M St., NW
Washington, DC 20036

1 Dr. D. J. Weiss
Dept. of Psychology
N660 Elliott Hall
University of Minnesota
Minneapolis, MN 55455

1 Dr. Keith Wescourt
Dept. of Psychology
Stanford University
Stanford, CA 94305

1 Dr. Earl Hunt
Dept. of Psychology
University of Washington
Seattle, WA 98105

1 Dr. Thomas G. Sticht
Assoc. Director, Basic Skills
National Institute of Education
1200 19th St., NW
Washington, DC 20208

1 Prof. Fumiko Samejima
Dept. of Psychology
Austin Peay Hall 304C
University of Tennessee
Knoxville, TN 37916